EXHIBIT R-1

Prepared: April 10, 2022

University of California, San Francisco

CURRICULUM VITAE

Name: John T Mongan, MD, PhD

Position: Associate Professor of Clinical Radiology, Step 2
Radiology
School of Medicine

Associate Chair, Translational Informatics
Director, Center for Intelligent Imaging

Address: Box 0628
University of California, San Francisco
Voice: 415-514-6002
Email: john.mongan@ucsf.edu

EDUCATION

1995 - 1999	Stanford University	B.S.	Chemistry with Honors and Distinction	
2000 - 2008	University of California San Diego	MD		
2002 - 2006	University of California San Diego	PhD	Bioinformatics	J. Andrew McCammon
2008 - 2009	Kaiser Oakland Medical Center	Intern	Internal Medicine	
2009 - 2013	University of California San Francisco	Resident	Diagnostic Radiology	
2013 - 2014	University of California San Francisco	Fellow	Ultrasound and Abdominal Imaging, Diagnostic Radiology	

LICENSES, CERTIFICATION

2009	California Medical License
2013	Diagnostic Radiology, American Board of Radiology
2013	California Radiology X-Ray Supervisor and Operator Permit
2016	Clinical Informatics, American Board of Preventive Medicine

PRINCIPAL POSITIONS HELD

2013 - 2014	University of California, San Francisco	Clinical Instructor	Radiology and Biomedical Imaging
2014 - 2019	University of California, San Francisco	Assistant Professor in Residence	Radiology and Biomedical Imaging
2019 - present	University of California, San Francisco	Associate Professor of Clinical Radiology	Radiology and Biomedical Imaging
2015 - 2016	University of California, San Francisco	Associate Chair, Informatics	Radiology and Biomedical Imaging
2016 - 2020	University of California, San Francisco	Vice Chair, Informatics	Radiology and Biomedical Imaging
2020 - present	University of California, San Francisco	Associate Chair, Translational Informatics	Radiology and Biomedical Imaging

OTHER POSITIONS HELD CONCURRENTLY

2013 - 2014	University of California, San Francisco	Chief Fellow, Non-ACGME Radiology Fellows	Radiology and Biomedical Imaging
2014 - present	VA Medical Center, San Francisco	Attending Radiologist	Radiology

HONORS AND AWARDS

1996	President's Award for Academic Excellence (top 3% of freshman class)	Stanford University
1997	Hoefer Prize for best undergraduate natural sciences writing	Stanford University
1998	Chemistry Department Analytical Chemistry Award	Stanford University
1999	Chemistry Department Marsden Award for top Chemistry graduate	Stanford University
2000	NIH Medical Scientist Training Program	University of California San Diego
2003	Taft Family Physical Sciences Fellow	University of California San Diego
2004	La Jolla Interfaces in Science Predoctoral Fellow	Burroughs Wellcome

2005	COMP division CCG Graduate Research Excellence Award	American Chemical Society
2005	Student Research Achievement Award	The Biophysical Society
2008	Free Clinic Leadership Award	University of California San Diego
2008	Medical School Merck Award for Outstanding Academic Accomplishments (top 3 graduates)	University of California San Diego
2013	Margulis Society Outstanding Radiology Resident Researcher	University of California San Francisco
2021	Editor's Recognition Award (recognizing consistent excellence as a reviewer)	RadioGraphics, Radiological Society of North America
2022	Fellow of the Society of Abdominal Radiology	Society of Abdominal Radiology

KEYWORDS/AREAS OF INTEREST

Informatics, artificial intelligence, deep learning, machine learning, electronic medical records, imaging comparative effectiveness research, contrast media

CLINICAL ACTIVITIES

CLINICAL ACTIVITIES SUMMARY

I was pleased to have my excellence as an Abdominal Radiologist and contributions to the field recognized this year by becoming a **Fellow of the Society of Abdominal Radiology**.

I am the clinical attending physician on the Abdominal Imaging and Ultrasound service **2 days per week 12 months per year**. On a typical day I am responsible for direct supervision of 2-4 medical students, residents and/or fellows. Approximately 30% of my clinical days are on services that perform biopsies; on these days in addition to reading imaging studies I typically teach and supervise or perform 4 image-guided biopsy procedures.

In addition to my service as a Radiologist, I continue to work as a **Clinical Informaticist** under my ABMS board certification in Clinical Informatics. My clinical informatics service constitutes **1.5 days per week 12 months per year.** This certification formalizes and recognizes the clinical judgment I routinely exercise in my leadership of the Imaging IT group (which provides PACS services) and the Radiology Innovation & Analytics team, through my role as Associate Chair, Translational Informatics. As these groups consist of non-physicians, my leadership brings the patient and physician perspectives and priorities to these groups, guiding their work to most effectively support our healthcare mission.

Organ Transplant is one of the marquee programs at UCSF; my ultrasound and abdominal imaging service is an integral component of this program. I perform pre-operative imaging evaluation of both living organ donors and recipients. The serial post-operative ultrasounds that I read are a primary assessment of transplanted organ health. Approximately 30% of my procedures are renal transplant biopsies. Additionally, I am often called upon by Transplant

Surgery to come to the operating room and assess the vascularity of organ grafts. My evaluation of the transplanted organ and its vascular anastamoses determine whether the patient will be closed as-is or will require immediate revision of the anastamoses.

I am the Radiology attending presenting imaging for several **tumor boards**. Each of these serves as both a clinical decision making and teaching conference, and involves the participation of attendings, fellows, residents and occasionally medical students from multiple departments. I present at these conferences on a rotating basis with my colleagues in the Ultrasound and Abdominal Imaging Sections; I present on average 2 of these conferences per month, which involves review of 10 to 30 imaging studies prior to the 1 to 1.5 hour conference. These conferences include:

- Liver Tumor Board (twice weekly): Attended by Interventional Radiology, Gastroenterology, Oncology, and Transplant Surgery
- Gastrointestinal Tumor Board (weekly): Attended by Pathology, Gastroenterology and Oncology
- Genito-Urinary Tumor Board (twice monthly): Attended by Pathology, Urology and Oncology
- Adrenal Tumor Board (monthly): Attended by Pathology, Endocrinology, and Surgery
- Gynecologic Oncology Tumor Board (weekly): Attended by Pathology, Medical Oncology and Gynecology Oncology

CLINICAL SERVICES

2014 - 2015	UCSF Abdominal and Ultrasound Radiology Sections, Attending	4 days per week, 12 months per year
2015 - 2019	UCSF Abdominal and Ultrasound Radiology Sections, Attending	2 days per week, 12 months per year
2019 - 2020	UCSF Abdominal and Ultrasound Radiology Sections, Attending	2.5 days per week, 12 months per year
2020 - present	UCSF Abdominal and Ultrasound Radiology Sections, Attending	2 days per week, 12 months per year

PROFESSIONAL ACTIVITIES

MEMBERSHIPS

2008 - present	Radiological Society of North America
2009 - 2017	American College of Radiology
2014 - 2016	Society of Radiologists in Ultrasound
2014 - present	Society of Abdominal Radiologists
2015 - present	Society for Imaging Informatics in Medicine

SERVICE TO PROFESSIONAL ORGANIZATIONS

2015 - 2021	Radiological Society of North America	RadLex Steering Committee Member
2018 - 2020	Radiological Society of North America	R&E Foundation Education Study Section Member
2018 - 2020	Radiological Society of North America	Scientific Program Committee Member
2018 - 2019	Radiological Society of North America	Machine Learning Steering Committee Member
2019 - 2020	Radiological Society of North America	Machine Learning Steering Committee, Vice Chair
2020 - present	Radiological Society of North America	Machine Learning Steering Committee, Chair
2020 - present	Society of Abdominal Radiology	Informatics Committee Member
2020 - present	Society of Abdominal Radiology	Artificial Intelligence Emerging Technology Committee Member
2021 - present	Radiological Society of North America	Imaging Informatics Meeting Program Subcommittee Member
2021 - present	Radiological Society of North America	Organizer, AI Safety Summit

SERVICE TO PROFESSIONAL PUBLICATIONS

2015 - present	Emergency Medicine Journal - Reviewer
2017 - present	Radiology - Reviewer
2018 - present	Radiology: Artificial Intelligence - Associate Editor
2018 - present	Radiographics - Informatics review panel
2019 - present	Journal of Digital Imaging - Reviewer
2021 - present	Journal of the American Medical Informatics Association - Reviewer
2021 - 2022	Radiology: Artificial Intelligence - Special issue Guest Editor
2021 - present	European Radiology - Reviewer

2021 - present	IEEE Transactions on Artificial Intelligence - Reviewer

INVITED PRESENTATIONS - INTERNATIONAL

2013	Dual Contrast Dual-Energy CT in Imaging of Penetrating Abdominal Trauma. Contrast Media Research Symposium. Beijing, China	
2017	Dose Reduction in Abdominal CT. Partnership for Dose Collaborative, Sausalito, California.	
2020	Radiology and Machine Learning: from Dialogue to Clinical Practice. Medical Image Computing and Computer Assisted Interventions (MICCAI) Annual Meeting. Scheduled for Lima, Peru; virtual due to pandemic.	Plenary session panel
2021	Critical Evaluation of AI for Purchase and Deployment. 102nd German Röntgen Congress (virtual due to pandemic).	Plenary session

INVITED PRESENTATIONS - NATIONAL

2004	Recent Developments in Constant pH Molecular Dynamics. AMBER Development Conference Stony Brook, New York. (AMBER is a leading software package for computational molecular dynamics simulations.)	
2005	Implementation of Molecular Surface Generalized Born. AMBER Development Conference, Salt Lake City, Utah.	
2012	Trauma Imaging with Color Contrast for Color CT: In Vivo Use of Complementary Contrast Materials at Dual-energy Computed Tomography. Radiological Society of North America, Chicago, Illinois	
2015	CT and MR Safety. UCSF Radiology Board Review CME Course, San Francisco, California	
2015	Ultrasound Technical Tips and Tricks. UCSF Radiology Board Review CME Course, San Francisco, California	
2015	Radiology Informatics. UCSF Radiology Fall Highlights CME course, San Francisco, California	
2016	CT and MR Safety. UCSF Radiology Board Review CME Course, San Francisco, California	
2016	Ultrasound Technical Tips and Tricks. UCSF Radiology Board Review CME Course, San Francisco, California	
2016	Thyroid Nodules: What to Evaluate, Which to Biopsy and How to Be Successful Every Time. Society of Abdominal Radiology, Waikoloa Village, Hawaii.	Workshop

2016	Context Integration of EMR (Electronic Medical Record) with PACS Increases Radiologist Use of EMR. Society of Imaging Informatics in Medicine, Portland, Oregon.	
2017	Ultrasound Technical Tips and Tricks. UCSF Radiology Board Review CME Course, San Francisco, California	
2017	CT and MR Safety. UCSF Radiology Board Review CME Course, San Francisco, California	
2017	Dual-energy CT Image Visualization. Society of Abdominal Radiology, Hollywood, Florida.	Workshop
2018	Ultrasound Technical Tips and Tricks. UCSF Radiology Board Review CME Course, San Francisco, California	
2018	CT and MR Safety. UCSF Radiology Board Review CME Course, San Francisco, California	
2018	Avoiding Pitfalls in Study Design and Data Analysis. Society of Abdominal Radiology, Scottsdale, Arizona.	
2018	Leveraging CDS Technology Successfully. Radiology Business Management Association 2018 PaRADigm, San Diego, California.	Plenary session panel
2018	Pancreatic Adenocarcinoma: Diagnosis, Staging and Mimics. UCSF Diagnostic Imaging Update, Kauai, Hawaii.	
2018	Understanding and Optimizing CT Radiation Dose. UCSF Diagnostic Imaging Update, Kauai, Hawaii.	
2018	Artificial Intelligence/Deep Learning: Implications for Radiologists. UCSF Diagnostic Imaging Update, Kauai, Hawaii.	
2018	What You Need to Know About the New Medicare Radiology Decision Support Requirement. UCSF Diagnostic Imaging Update, Kauai, Hawaii.	
2018	How to be Successful Every Time with US-Guided Thyroid Biopsy. UCSF Diagnostic Imaging Update, Kauai, Hawaii.	
2018	Ultrasound Evaluation of Organ Transplants. UCSF Diagnostic Imaging Update, Kauai, Hawaii.	
2018	Artificial Intelligence in Radiology. Radiological Society of North America, Chicago, Illinois.	Moderator
2019	Dataset Curation and Annotation. American Roentgen Ray Society, Honolulu, Hawaii.	

2019	Lifecycle of a Radiology Exam. National Imaging Informatics Course, hosted online by Radiological Society of North America and Society for Imaging Informatics in Medicine.	Plenary speaker
2019	Ultrasound Technical Tips and Tricks. UCSF Radiology Board Review CME Course. San Francisco, California	
2019	Playing it Safe: What you Need to Know About MR and CT Safety. UCSF Radiology Board Review CME Course, San Francisco, California	
2019	How To Be Successful with US Thyroid Biopsy. UCSF Diagnostic Imaging Update, Maui, Hawaii.	
2019	US of Organ Transplants. UCSF Diagnostic Imaging Update, Maui, Hawaii.	
2019	Improved US Diagnosis: Technical Tips. UCSF Diagnostic Imaging Update, Maui, Hawaii.	
2019	Pancreatic adenocarcinoma: Diagnosis, Staging, Mimics. UCSF Diagnostic Imaging Update, Maui, Hawaii.	
2019	CT/MR Safety. UCSF Diagnostic Imaging Update, Maui, Hawaii.	
2019	AI and Deep Learning: Implications for Radiologists. UCSF Diagnostic Imaging Update, Maui, Hawaii.	
2019	Artificial Intelligence: NLP and Reporting. Radiological Society of North America, Chicago, Illinois.	Moderator
2020	Thyroid Nodules: What To Evaluate, Which To Biopsy and How To Be Successful Every Time. Society of Abdominal Radiology, Maui, Hawaii.	
2020	Sonographic Evaluation of Kidney and Liver Transplants. Association of Program Directors in Radiology, National Noon Conference Series. (virtual for Covid-19)	
2020	Decision Support and Implications for Federal Regulations (PAMA): What You Need to Know and Do. Radiological Society of North America, Chicago, Illinois. (virtual for Covid-19)	Panelist
2020	Pulmonary Embolism Machine Learning Challenge Winners and Awards. Radiological Society of North America, Chicago, Illinois. (virtual for Covid-19)	Panelist
2021	Ultrasound: Technical Tips & Artifacts. UCSF Radiology Annual Review, San Francisco, California (virtual for Covid-19)	

2021	Radiology Safety: Need to Know. UCSF Radiology Annual Review, San Francisco, California (virtual for Covid-19)	
2021	Automated detection of IVC Filters with a Deep Object Detection Network. Society for Imaging Informatics in Medicine (virtual for Covid-19)	Podium
2021	Behind the scenes of the 2021 SIIM-FISABIO-RSNA Covid-19 Pneumonia Detection Challenge. Society for Imaging Informatics in Medicine (virtual for Covid-19)	Panelist
2021	Imaging CDS, AUC, and Lessons Learned. Healthcare Information and Management Systems Society (virtual for Covid-19)	Panelist
2021	SIIM-FISABIO-RSNA Covid-19 Pneumonia Detection Kaggle Challenge Winners' Showcase. Conference on Machine Intelligence in Medical Imaging (virtual for Covid-19)	Panelist
2021	Safe and Effective Translation of Artificial Intelligence: From the Lab to the Reading Room. University of Washington, Seattle, Washington.	
2021	Current State of AI in Radiology. Radiological Society of North America, Chicago, Illinois.	Panelist
2021	Medical Imaging and Data Resource Center: A Multi-Society Approach to Advance Research on COVID-19. Radiological Society of North America, Chicago, Illinois.	Panelist
2021	Failure in Clinical Artificial Intelligence: Lessons from Aviation. Radiological Society of North America, Chicago, Illinois.	
2021	Best Practices in Radiology: Clinical Decision Support Rollout. Radiological Society of North America, Chicago, Illinois.	Panelist
2021	RSNA AI Challenge: Brain Tumor AI Challenge Recognition Event. Radiological Society of North America, Chicago, Illinois.	Panelist
2021	Data Normalization. RSNA Imaging Artificial Intelligence Certificate Program (virtual)	
2021	Data Annotation. RSNA Imaging Artificial Intelligence Certificate Program (virtual)	
2022	Pancreatic Adenocarcinoma: How to Diagnose, Stage and Recognize Mimics. UCSF Body Imaging: Abdominal & Thoracic. Kona, Hawaii.	

2022	Artificial Intelligence: What Does it Mean for Radiologists? UCSF Body Imaging: Abdominal & Thoracic. Kona, Hawaii.	
2022	Artificial Intelligence: How to Evaluate and Purchase. UCSF Body Imaging: Abdominal & Thoracic. Kona, Hawaii.	
2022	Ultrasound Evaluation of Organ Transplants. UCSF Body Imaging: Abdominal & Thoracic. Kona, Hawaii.	
2022	US-guided Thyroid Biopsy: How to be Successful Every Time. UCSF Body Imaging: Abdominal & Thoracic. Kona, Hawaii.	
2022	New Medicare Radiology Decision Support Requirement: What You Need to Know. UCSF Body Imaging: Abdominal & Thoracic. Kona, Hawaii.	
2022	Purchasing AI Applications. Society of Abdominal Radiology, Scottsdale, Arizona.	Plenary speaker
2022	Critical Evaluation of AI for Purchase and Deployment. Society of Abdominal Radiology, Scottsdale, Arizona.	

INVITED PRESENTATIONS - REGIONAL AND OTHER INVITED PRESENTATIONS

2005	Fast Molecular Surface Generalized Born. Accelrys Corporation San Diego, California.
2014	Using Radiology Effectively. UCSF Internal Medicine Residents.
2014	Ultrasound-guided FNA: How to hit the lesion (almost) every time. UCSF Radiology Residents.
2014	Using Radiology Effectively. UCSF Emergency Medicine Resident Conference and CME Lecture.
2015	CT Systems. UCSF Radiology Residents.
2015	CT Dosimetry. UCSF Radiology Residents.
2015	Ultrasound of Transplants. UCSF Radiology Sonography for Sonographers CME course.
2015	Radiology Physics Review. UCSF Radiology Residents.
2015	Ultrasound of solid organ transplants. UCSF Radiology Residents.
2015	Statistics in R for Radiologists. UCSF Radiology Residents.
2015	PAMA and Clinical Decision Support. UCSF Radiology Residents.

2015	Pancreatic Adenocarcinoma and Mimics. UCSF Abdominal Imaging Fellows.	
2016	Natural Language Processing in Radiology. UCSF Radiology Residents.	
2016	Clinical Decision Support: Realizing the Potential of Computers in Medicine. UC Davis Radiology Grand Rounds, Sacramento, California	Grand Rounds
2016	Ultrasound-guided Thyroid Fine Needle Aspiration. UC Davis Radiology Residents, Sacramento, California	
2016	CT Systems. UCSF Radiology Residents.	
2016	CT Dosimetry. UCSF Radiology Residents.	
2016	Ultrasound-guided Thryoid Fine Needle Aspiration. UCSF Abdominal Imaging Fellows.	
2016	Clinical Decision Support: Realizing the Potential of Computers in Medicine. UCSF Radiology Residents.	
2016	UCSF Radiology Informatics Overview. UCSF Radiology Residents.	
2016	Ultrasound-guided Thyroid Fine Needle Aspiration. UCSF Radiology Residents.	
2017	Using Radiology Effectively. UCSF Internal Medicine Residents.	
2017	A Celebration of 25 Years of Radiological Informatics. UCSF Radiology Grand Rounds.	Grand Rounds
2017	Ultrasound Evaluation of Solid Organ Transplants. UCSF Radiology Residents.	
2017	Using Radiology Effectively. UCSF Internal Medicine Residents.	
2017	Ultrasound-guided Thyroid Fine Needle Aspiration. UCSF Radiology Residents.	
2017	Using Radiology Effectively. OneMedical Grand Rounds, San Francisco, California.	Grand Rounds
2017	Artificial Intelligence and the Future of Radiology. UCSF Osher Mini-Med School.	
2018	CT Systems. UCSF Radiology Residents.	
2018	CT Dosimetry. UCSF Radiology Residents.	
2018	Artificial Intelligence in Medicine. UCSF Division of Neuroinflammation and Glial Biology.	Grand Rounds

2018	Ultrasound Evaluation of Solid Organ Transplants. UCSF Radiology Residents.	
2018	Pancreatic Adenocarcinoma and Mimics. UCSF Radiology Residents.	
2018	Ultrasound Evaluation of Solid Organ Transplants. UCSF Abdominal Imaging Fellows.	
2018	Artificial Intelligence in Radiology. UCSF Radiology Residents.	
2019	Ultrasound Evaluation of Solid Organ Transplants. UCSF Abdominal Imaging Fellows.	
2019	Ultrasound-guided Thyroid Fine Needle Aspiration. UCSF Abdominal Imaging Fellows.	
2019	Artificial Intelligence in Medicine. Fromm Institute, San Francisco, California.	
2019	Artificial Intelligence: What is it and what does it mean for radiologists. UCSF Radiology Residents.	
2020	Ultrasound Evaluation of Solid Organ Transplants. UCSF Radiology Residents.	
2020	The Role of Health Informatics Professionals in Medical AI. UCSF Health Informatics	Grand Rounds
2020	Ultrasound-guided Thyroid Fine Needle Aspiration. UCSF Abdominal Imaging Fellows.	
2020	Ultrasound Evaluation of Solid Organ Transplants. UCSF Abdominal Imaging Fellows.	
2020	CT Systems. UCSF Radiology Residents.	
2021	Ultrasound Physics. UCSF Radiology Residents.	
2021	Ultrasound MCQ Board Review. UCSF Radiology Residents.	
2021	CT Dosimetry. UCSF Radiology Residents.	
2021	Pancreatic Adenocarcinoma Diagnosis, Staging and Mimics. UCSF Abdominal Imaging Fellows.	
2021	Ultrasound Evaluation of Solid Organ Transplants. UCSF Abdominal Imaging Fellows.	
2021	Ultrasound-guided Thyroid Fine Needle Aspiration. UCSF Abdominal Imaging Fellows.	
2021	Ultrasound Evaluation of Solid Organ Transplants. UCSF Radiology Residents.	

2021	CT Systems. UCSF Radiology Residents.	
2022	Ultrasound Physics. UCSF Radiology Residents.	

GOVERNMENT AND OTHER PROFESSIONAL SERVICE

2020 - 2021	DFG (German Research Foundation)	Artificial Intelligence Grant Reviewer
2021 - 2021	STARD-AI reporting guideline	Expert Consensus Panel Member

UNIVERSITY AND PUBLIC SERVICE

SERVICE ACTIVITIES SUMMARY

Since submission of my CV in spring 2020 for my most recent advancement, effective 7/1/21, I have successfully **led the replacement of the decades-old Radiology PACS, led the deployment of the first artificial intelligence (AI) algorithm into regular clinical use by Radiology and led the renewal of University of California-wide certification by the Center for Medicare & Medicaid Services (CMS)** for creation of imaging Appropriate Use Criteria. My campus leadership in informatics and IT has been recognized by **appointment to chair the Chief Research Informatics Officer (CRIO) Task Force** and serve on the Associate Chief Informatics Officer - Research interview panel.

PACS is the software used by radiologists to display and manipulate medical images. It is used all day, every day by every UCSF radiologist working clinically, and is essential to the efficiency of clinical operation of the department. Building on my success in representing UCSF in the UC-wide PACS Selection RFP, in partnership with the Associate Chair Clinical Informatics over the last two years I **led the successful replacement of the decades-old PACS system**, replacing four poorly integrated clinical viewers with a single, consolidated clinical viewer software (Visage). This project supported 10% of my effort during calendar year 2021, including assembling and co-chairing a steering committee consisting of representatives from each clinical section in Radiology, hiring and coordinating a team of consultants and meeting regularly with vendors and the Imaging IT team. We **completed the $10.3M project on time and under budget, delivering a system that radiologists are enthusiastic about** using and found more efficient than the old system even on their first day using it. This is the first Tier 1 clinical system at UCSF to be deployed in the cloud. Our successful go-live in December 2021 **served as the model for subsequent implementations at UCI and UCSD.** The success of the project and my leadership of it was recognized in the UCSF CIO news: https://it.ucsf.edu/news/ucsf-it-radiology-collaborate-solve-decade-long-challenge .

In my role as a **Director of the UCSF Center for Intelligent Imaging (Ci2), I led the deployment of UCSF's first imaging-based AI algorithm** into regular clinical usage. I had previously served as UCSF's site PI for the multi-center validation of the algorithm, which enables coronary artery calcium scoring, a strong predictor of coronary artery disease, to be performed on all non-contrast chest CTs, rather than only logistically complex cardiac-gated CTs. To guide this process, **I developed a framework for evaluating whether and how an AI algorithm should be deployed** clinically, which was used by the Ci2 Clinical Deployment committee that I chair to evaluate and approve the algorithm. The committee is currently evaluating two additional AI algorithms using the framework. In this role I also promote the

infrastructure, education and research missions of Ci2 through chairing the Scientific Computing Services Steering Committee and serving on the Infrastructure, Education and Scientific Research Group Committees.

As the **chair of the UC-wide Appropriate Imaging Coalition** (https://qple.ucop.edu) **I led the successful renewal of certification of the University of California by the Center for Medicare and Medicaid Services (CMS)** as a qualified provider-led entity (QPLE) under the imaging clinical decision report requirement of the Protecting Access to Medicare Act (PAMA). This makes UC one of only twelve healthcare organizations in the country certified to create Appropriate Use Criteria (AUC) to meet the requirement. The six AUC for which I led development through the Coalition continue to be valid due to this recertification; **without this recertification, the five medical campuses of UC would no longer be paid for CT, MR or nuclear medicine studies of Medicare patients** after January 1, 2023. As chair of the Coalition, my on-going responsibilities include supervising and coordinating six multidisciplinary teams in their annual review and revision of the AUC, supervision of a medical librarian employed by the program and negotiating, securing and managing the budget of $120k/yr. My responsibilities as chair of this program require substantial investment of time; I am **supported by the program at 10% effort.**

Compliance with CMS regulations requires **implementation of the AUC set into a clinical decision support system to be consulted for every outpatient Medicare imaging order**. I continue to lead the implementation and optimization of a compliant system at UCSF, in conjunction with physician representatives from Emergency and Ambulatory Medicine, as part of the **Imaging Clinical Decision Support Implementation Task Force**.

I employ my **expertise in IT and informatics to provide campus-wide service** in multiple venues. In the past year I was appointed **chair of the Chief Research Informatics Officer (CRIO) Task Force**, charged with determining whether UCSF should create a CRIO position, and served as a member of the Associate Chief Informatics Officer - Research interview panel to select UCSF's first ACIO - Research. I continue to serve on the campus-wide **Enterprise Imaging (VNA) Steering Committee**, providing guidance for the UCSF implementation and use of the archive. As a member of the **IT Governance Committee on Research Technology**, I provide perspective of the computational needs of both clinical researchers and imaging researchers.

UNIVERSITY SERVICE
UC SYSTEM AND MULTI-CAMPUS SERVICE

2015 - 2015	Imaging Ordering Clinical Decision Support Proposal Review Committee	Member
2015 - 2017	UC Enterprise Imaging Steering Committee	Member
2015 - present	University of California Imaging Appropriate Use Criteria (QPLE) Steering Committee	Chair
2017 - 2020	UC Imaging Decision Support Steering Committee	Member
2018 - 2020	UC PACS Selection Committee	Member

UCSF CAMPUSWIDE

2014 - 2018	Resource Allocation Program (RAP) Grant Review Committee	Member
2016 - 2017	Institute for Computational Health Sciences (ICHS) Faculty Search Committee (JPF01218)	Member
2017 - 2018	Enterprise Cloud Service Working Group	Member
2018 - present	Enterprise Imaging (VNA) Steering Committee	Member
2018 - present	IT Governance Committee on Research Technology	Member
2019 - present	Imaging Clinical Decision Support (CDS) Implementation Task Force	Member
2020 - present	Center for Intelligent Imaging (CI2) Steering Committee	Member
2020 - present	Center for Intelligent Imaging Clinical Deployment Pillar Committee	Chair
2020 - present	Center for Intelligent Imaging Education Pillar Committee	Member
2020 - present	Center for Intelligent Imaging Scientific Research Group Committee	Member
2020 - present	Center for Intelligent Imaging Infrastructure Committee	Member
2022 - present	Chief Research Informatics Officer Task Force	Chair
2022 - present	Associate Chief Information Officer Research Interview Panel	Member

SCHOOL OF MEDICINE

2011 - 2014	APeX (Epic Electronic Medical Record) Fellows and Residents Advisory Group	Member
2017 - 2017	Strategic Plan Working Group: Data and Technology	Member

DEPARTMENTAL SERVICE

2012 - 2013	Resident Quality Improvement Project (Radiation Dose Reporting)	Project Leader
2013 - 2014	Non-ACGME Radiology Fellows	Chief Fellow
2015 - present	Operations Committee	Member
2015 - present	Executive Committee	Member
2019 - present	Scientific Computing Services Steering Committee	Chair
2021 - present	AIIMS (New PACS Deployment) Steering Committee	Co-chair

SERVICE AT OTHER UNIVERSITIES

2020 - present	CSU Chico Cybersecurity Advisory Board	Member

COMMUNITY AND PUBLIC SERVICE

2016 - present Bay Area Science Festival — UCSF Radiology Volunteer

CONTRIBUTIONS TO DIVERSITY

CONTRIBUTIONS TO DIVERSITY Contributions to Diversity, Equity & Inclusion Guidance

Since submission of my CV in spring of 2020 for my most recent advancement, effective 7/1/21, I have continued to build on my Differences Matter Diversity, Equity and Inclusion Champion training by **mentoring and sponsoring two additional women in the faculty and residency** in informatics and artificial intelligence activities. Artificial intelligence (AI) and informatics suffer from even lower representation of women and under-represented minorities than radiology as a whole, and I have been working to address this on an individual basis through identifying under-represented people with interests in these areas, mentoring them and promoting and sponsoring their involvement in related activities. I have begun mentoring Katie Grouse (neurology faculty) in research, involving her in a sponsored research project with Siemens to add AI understanding of text-based reason for exam to the clinical decision support system used for ordering imaging. I also began mentorship of Maggie Chung, one of our radiology residents, in AI research, which has led to a joint conference abstract and manuscript currently under revision for the journal Radiology. I continue to sponsor and mentor Kim Kallianos' work in AI, partnering with her to deploy into routine clinical practice the AI-based coronary calcium scoring algorithm that we worked on together.

I mentored Igor Teodoro, a Brazilian medical student with an interest in informatics and clinical decision support, involving him in the work of the UC Appropriate Imaging Coalition that I chair.

I have mentored and sponsored of Tatiana Kelil, including promoting her as our department's expert on 3D printing, nominating her to take over my position on the UCSF RAP grant review committee and nominating her for a position on the Informatics section of the Radiological Society of North America Scientific Program Committee.

I have begun a new research project, SYRMOUNT, which applies AI to retrospectively analyze the reasons for and outcomes of imaging in women with breast cancer with a goal of identifying how imaging can be used more effectively for these women.

TEACHING AND MENTORING

TEACHING SUMMARY

Since submission of my CV in spring 2020 for my most recent advancement, effective 7/1/21, I have leveraged my expertise in AI to serve as a lecturer for the Masters of Science in Biomedical Imaging (MSBI program) through BioEng 245. I have continued to play a major role in CME efforts, serving as a primary lecturer (delivering 6 lectures) at a recent UCSF Radiology CME course.

My continuing teaching activities include:

- **Informal Teaching** - Most of my teaching time is in the setting of the reading room. Approximately 95% of my clinical work is done in conjunction with a resident or fellow. For diagnostic studies, this involves sitting side by side, reviewing each imaging finding as I read each case, and discussing how to put the findings into the context of the patient history to arrive at an imaging differential diagnosis and impression. For biopsies, I discuss the steps of the procedure and approach with trainees immediately before we begin. I directly supervise my trainees in each procedure, reviewing their planned approach, providing guidance during the procedure, and stepping in to perform portions of the procedure when necessary. I always

provide a debrief session after each procedure to discuss what went well, what needed improvement and how the next biopsy can be better. When we have medical students in the reading room, I actively engage them and get them logged in to their own imaging station (PACS) so they can actively participate in reading imaging. I tailor my instruction of them towards what will have greatest utility to them in their future specialty within medicine. Clinical teams frequently consult us in the reading room; interacting with these teams gives me an opportunity to teach imaging findings and effective use of imaging to trainees from other departments, as well as to learn from them what questions and pieces of information are most important to them.

- **Formal Teaching** - Drawing on the combination of my experience in AI and clinical Radiology, I lecture in BioEng 245 on the keys to developing and deploying clinically effective AI. I find formal teaching of students highly rewarding and welcome the opportunity to share what I love about my research and the specialty of medicine to which I've devoted my career.
- **Lecturing Outside of Scheduled Classes** - I regularly lecture radiology trainees at our daily morning and noon conferences. Drawing on my PhD work in computational biophysics, many of my lectures fall within the radiology physics curriculum. Beyond the radiology department, I make an active effort to engage residents from other programs and have delivered lectures on effective use of imaging to both the internal medicine and emergency medicine residency programs.

FORMAL TEACHING

	Academic Yr	Course No. & Title	Teaching Contribution	School	Class Size
	2013 - 2014	140.03 Diagnostic Radiology	Ultrasound lab instructor	Medicine	9
	2014 - 2015	140.03 Diagnostic Radiology	Ultrasound lab instructor	Medicine	14
	2014 - 2015	140.03 Diagnostic Radiology	Ultrasound lab instructor	Medicine	15
	2014 - 2015	140.03 Diagnostic Radiology	Ultrasound lab instructor	Medicine	10
	2014 - 2015	140.03 Diagnostic Radiology	Ultrasound lab instructor	Medicine	23
	2014 - 2015	140.03 Diagnostic Radiology	Ultrasound lab instructor	Medicine	12
	2014 - 2015	140.03 Diagnostic Radiology	Ultrasound lab instructor	Medicine	12
	2015 - 2016	140.03 Diagnostic Radiology	Ultrasound lab instructor	Medicine	15
	2015 - 2016	140.03 Diagnostic Radiology	Ultrasound lab instructor	Medicine	17
	2015 - 2016	140.03 Diagnostic Radiology	Ultrasound lab instructor	Medicine	17

	Academic Yr	Course No. & Title	Teaching Contribution	School	Class Size
	2015 - 2016	140.03 Diagnostic Radiology	Ultrasound lab instructor	Medicine	14
	2015 - 2016	140.03 Diagnostic Radiology	Ultrasound lab instructor	Medicine	16
	2015 - 2016	198 Independent Study in Radiology	Mentor	Medicine	1
	2015 - 2016	170.07 Current Issues in Medical Informatics	Guest lecturer	Medicine	3
	2016 - 2017	IDS 106: Methods, Mechanisms and Malignancies	Small group instructor	Medicine	12
	2016 - 2017	RAD 140.19: Advanced Clinical Clerkship in Radiology	Supervising radiologist	Medicine	1
	2016 - 2017	RAD 140.03: Diagnostic Radiology	Ultrasound lab instructor	Medicine	17
	2016 - 2017	RAD 140.03: Diagnostic Radiology	Ultrasound lab instructor	Medicine	14
	2016 - 2017	RAD 140.03: Diagnostic Radiology	Ultrasound lab instructor	Medicine	15
	2016 - 2017	RAD 140.03: Diagnostic Radiology	Ultrasound lab instructor	Medicine	15
	2016 - 2017	RAD 140.03: Diagnostic Radiology	Ultrasound lab instructor	Medicine	17
	2017 - 2018	RAD 140.03: Diagnostic Radiology	Ultrasound lab instructor	Medicine	15
	2017 - 2018	RAD 140.03: Diagnostic Radiology	Ultrasound lab instructor	Medicine	15
	2017 - 2018	RAD 140.03: Diagnostic Radiology	Ultrasound lab instructor	Medicine	18
	2017 - 2018	RAD 140.03: Diagnostic Radiology	Ultrasound lab instructor	Medicine	19
	2017 - 2018	IDS 113: Appropriate Use of Diagnostic Test	Lecturer	Medicine	20-50

	Academic Yr	Course No. & Title	Teaching Contribution	School	Class Size
	2017 - 2018	RAD 140.03: Diagnostic Radiology	Ultrasound lab instructor	Medicine	24
	2018 - 2019	IDS 113: Appropriate Use of Diagnostic Test	Lecturer	Medicine	20-50
	2018 - 2019	RAD 170.04: Intro to Radiology	Lecturer	Medicine	18
	2019 - 2020	RAD 140.19: Advanced Clinical Clerkship in Radiology	Supervising Radiologist	Medicine	1
	2020 - 2021	BioEng 245: Machine Learning Algorithms for Medical Imaging	Lecturer	Grad	12

INFORMAL TEACHING

2014 - present Residents and Fellows: I am on clinical service 2 days per week, 12 months per year; every one of these days I have 2-4 residents and/or fellows under my direct supervision. Approximately 85% of my time spent reading cases is in direct interaction with a trainee. Each case presents a teaching opportunity; I discuss the findings and their meaning with the trainee as I read each case.

2014 - present Medical Students: Students frequently spend an extended period of time in the reading room as part of RAD 140.19 "Advanced Clinical Clerkship in Radiology." I engage these students and get them logged into their own computer displaying images so they can start doing radiology. I tailor their activities and my instruction to their goals: for future radiologists, I help to start them down the path of being able to interpret studies independently; for future non-radiology clinicians, I focus on how to use radiology effectively and how to understand the meaning and utility of imaging diagnoses.

MENTORING SUMMARY

Since submission of my CV in spring 2020 for my most recent advancement, effective 7/1/21, **I have mentored six faculty members and two trainees**:

- Katie Grouse is an Asst. Professor of Neurology at UCSF. She works in clinical informatics as part of CHIO Russ Cucina's team. I provide research mentorship in the areas of clinical decision support and artificial intelligence. **This led to shared research support from Siemens.**
- Kim Kallianos is an Asst. Professor of Radiology at UCSF. I provide scholarly and research mentorship on how to plan and undertake studies in artificial intelligence. **This led to joint publications in Nature - Digital Medicine and Clinical Radiology, shared research support from GE, and partnership in clinical deployment of an AI algorithm.**

- Andrew Taylor is an Assoc. Professor of Radiology at UCSF who decided to change his research focus to informatics in 2016. I provide career mentorship to him as he develops his skills and position as a leader in radiology informatics as well as research mentorship on how to structure informatics research. **This led to a joint publication in PLoS: Medicine and shared research support from GE.**
- Tatiana Kelil is an Asst. Professor of Radiology at UCSF. I provide career mentoring to her centering on her role in informatics as well as scholarly mentorship concerning developing an academic program around her 3D expertise.
- Aaron Kornblith is an Assoc. Professor of Emergency Medicine at UCSF. I provide research mentorship supporting his program of developing artificial intelligence-enabled ultrasound systems to reduce the need for CT radiation exposure in pediatric trauma victims. **This led to a K23 grant submission where I am formally listed as a mentor.**
- Roozbeh Houshyar is Director of Informatics and Assoc. Professor of Radiology at UCI. I provide career mentorship centering on his leadership role in radiology informatics within his department and his work implementing clinical systems including Epic/Radiant and PACS, and research mentorship on his work in artificial intelligence and publications in informatics. **This led to a joint publication in Emergency Radiology.**
- Kirti Maguida was a T-32 research and clinical fellow. I provided mentorship on her research projects in artificial intelligence (focusing on prostate) and career and clinical mentorship on her development as a clinical radiologist an informatics researcher.
- Maggie Chung is a radiology resident who will be a breast imaging fellow next year. I mentor her research projects in artificial intelligence and provide career advising. **This led to a joint conference abstract and a joint publication currently in revision for Radiology.**

PREDOCTORAL STUDENTS SUPERVISED OR MENTORED

Dates	Name	Program or School	Mentor Type	Role	Current Position
2011 - 2013	Samira Rathnayake	UCSF Medical Student	Research/Scholarly Mentor	Daily direct supervision and interaction 2011-12; interaction several times per month 2012-13.	Radiologist, Bay Imaging Consultants
2015 - 2016	Erik Velez	UCSF Medical Student	Research/Scholarly Mentor	Monthly in-person meetings; weekly email.	Radiology Resident, USC
2018 - 2019	Igor Teodoro	Brazilian Medical Student	Research/Scholarly Mentor,Career Mentor	Twice weekly in person meetings; guided readings; IT shadowing	Radiology Resident

POSTDOCTORAL FELLOWS AND RESIDENTS MENTORED

Dates	Name	Fellow	Mentor Role	Faculty Role	Current Position
2014 - 2015	Yi Li	Resident	Project Mentor	In person meetings monthly, email weekly	Asst. Prof., UCSF

Prepared: April 10, 2022

Dates	Name	Fellow	Mentor Role	Faculty Role	Current Position
2014 - 2017	Vignesh Arasu	Resident	Career Mentor	In person meetings 2-3 times per year	Radiologist at Kaiser, Investigator Kaiser Division of Reseach
2014 - 2015	Mark Kovacs	Clinical Fellow	Research/Scholarly Mentor	In person meetings monthly; email interaction weekly	Asst. Prof., Medical University of South Carolina
2015 - 2016	Bernice Lau	Clinical Fellow	Research/Scholarly Mentor	In person meetings monthly; email interaction weekly	Private Practice Radiologist, Calgary, Canada
2015 - 2017	Jenny Wan	Resident	Research/Scholarly Mentor	In-person meetings every 2-3 months; email interaction monthly	Private Practice Radiologist, Burlingame, CA
2015 - 2016	Matt Barkovich	Resident	Project Mentor	In-person meetings every 2-3 months; email interaction about twice monthly	Asst. Prof., UCSF
2017 - 2018	Tatiana Kelil	Clinical Felow	Research/Scholarly Mentor,Career Mentor	Email every month, telephone or in person meeting every 2-3 months	Asst. Prof., UCSF
2019 - 2020	Alex Chan	Clinical Fellow	Research/Scholarly Mentor,Project Mentor,Career Mentor	Email 2-3 times per month, in person meeting monthly	Asst. Prof, Mayo Clinic
2019 - 2020	Ravi Rajpoot	Clinical Fellow	Career Mentor	In person meetings twice monthly	Private Practice Radiologist

Dates	Name	Fellow	Mentor Role	Faculty Role	Current Position
2019 - 2021	Kirti Magudia	T-32 and Clinical Fellow	Research/Scholarly Mentor,Career Mentor,Co-Mentor/Clinical Mentor	Email 2-3 times per month, in person meeting every other month	Asst. Prof, Duke
2020 - present	Maggie Chung	Resident	Research/Scholarly Mentor,Career Mentor	Meetings twice monthly	UCSF Resident

FACULTY MENTORING

Dates	Name	Position while Mentored	Mentor Type	Mentoring Role	Current Position
2015 - 2019	Mark Kovacs	Asst. Professor	Research/Scholarly Mentor,Career Mentor	Email several times per month, in person meetings at conferences about twice per year. Produced two shared conference abstracts and a co-authored publication (submitted).	Asst. Prof., Medical University of South Carolina
2015 - present	Roozbeh Houshyar	Asst. Professor	Research/Scholarly Mentor,Career Mentor	Email approximately once per month, telephone calls about six times per year, in person meetings about twice per year.	Asst. Prof., UC Irvine
2016 - 2019	Tom Loehfelm	Asst. Professor	Career Mentor	Email 4-5 times per year, in person meetings twice per year	Asst. Prof., UC Davis
2016 - present	Andrew Taylor	Asst. and Assoc. Professor	Research/Scholarly Mentor,Career Mentor	Email several times per week, in person meetings every other week	Assoc. Prof., UCSF
2018 - 2020	Hailey Choi	Asst. Professor	Research/Scholarly Mentor,Career Mentor	Email several times per month, in person meetings monthly	Asst. Prof., UCSF

Dates	Name	Position while Mentored	Mentor Type	Mentoring Role	Current Position
2018 - present	Tatiana Kelil	Asst. Professor	Research/Scholarly Mentor,Career Mentor	Email monthly, meeting or phone call every 2-3 months	Asst. Prof., UCSF
2019 - present	Kim Kallianos	Asst. Professor	Research/Scholarly Mentor	Email weekly, in person meeting monthly	Asst. Prof, UCSF
2019 - present	Aaron Kornblith	Asst. and Assoc. Professor (Emergency Med)	Research/Scholarly Mentor	Email monthly, in person meeting quarterly	Assoc. Prof, UCSF
2020 - present	Katie Grouse	Asst. Professor (Neurology)	Research/Scholarly Mentor	Virtual meetings every other week	Asst. Prof, UCSF

RESEARCH AND CREATIVE ACTIVITIES

RESEARCH PROGRAM (SEPARATE SUMMARY)

Since submission of my CV in spring 2020 for my most recent advancement, effective 7/1/21, my research has focused primarily on **artificial intelligence (AI) (8 publications and 4 new grants and research support contracts)**.

Recognizing the central importance of data in machine learning forms of artificial intelligence, a major focus of my recent research has been on **assembly of large, well-curated and annotated, multi-institutional publicly available imaging datasets**. These include the RSNA International COVID-19 Open Annotated Radiology Database (RICORD), which subsequently formed the basis for the COVID-19 Medical Imaging and Data Resource Center; the RSNA Brain CT Hemorrhage Dataset; and the RSNA Pulmonary Embolism CT Dataset. I obtained grant funding to support my work on MIDRC as a whole, as well as additional grant funding to support the costs of contributing UCSF data to this project. Through my work as the chair of the RSNA Machine Learning Committee, each of these datasets has been promoted through large-scale public machine learning challenges with five-figure prizes hosted on Kaggle.

I am **bringing together my expertise in AI research with my background in imaging clinical decision support and appropriateness through two funded projects**. The first of these, SYRMOUNT, addresses a central challenge with development of evidence-based appropriate use criteria for imaging: there is very little evidentiary basis for most imaging. Datapoints as simple as the pretest probabilities for common imaging indications are almost entirely missing from the literature. In the SYRMOUNT project, I am leading the development of AI models to enable rapid analysis of very large retrospective datasets of imaging. Understanding why imaging was performed and what the imaging outcomes were will provide a major leap forward for evidence-based appropriate use of imaging. A second project uses AI to address the implementation of clinical decision support (CDS). At present, in order for the CDS software to understand why the study is being performed, the ordering provider must

select one or more indications from a list. This is generally less efficient than the traditional free-text reason for exam, and substantially limits the detail about the patient that can be provided to the radiologist. In this sponsored collaboration with Siemens, the UC-wide vendor for imaging CDS software, we are developing AI Natural Language Processing models to enable the CDS to understand the reason that a study is being performed from a free-text reason for exam.

RESEARCH AWARDS - CURRENT

1. 75N92020C00008	Project Co-PI	10 % effort	Geiger (PI)
NIBIB		08/21/2020	08/20/2022
Medical Imaging and Data Resource Center (MIDRC)		$ 3,000,000 direct/yr 1	$ 6,000,000 total

The MIDRC is a large project encompassing the RSNA, ACR and the AAPM to develop infrastructure for obtaining, curating, deidentifying and publicly releasing medical image data. The initial charge is create COVID-19 related datasets that support development of improved imaging-based diagnostic, prognostic and treatment guidance technologies, particularly those incorporating artificial intelligence

I am co-PI on two projects of this contract, one that seeks to develop a large scale repository of multi-institutional imaging and clinical data within the RSNA and a second that develops infrastructure for processing, normalizing and transferring the data from the RSNA repository into the central MIDRC resource where it can be made available to researchers. For both of these projects, I interface with data contributors and supervise and direct the work of RSNA staff in constructing and optimizing infrastructure.

2. 2021 DH Mongan C00239380	PI	1 % effort	Mongan (PI)
Siemens		09/10/2021	09/14/2022
Natural Language Processing of Reason for Exam in Imaging Clinical Decision Support		$ 37,340 direct/yr 1	$ 71,697 total

A major limitation of the Clinical Decision Support (CDS) software required by the Protecting Access to Medicare Act is that ordering providers must indicate the reason they are ordering imaging using a series of check boxes, in addition to the usual free text field, so that the software can understand why the imaging study is being requested. This structured indication capture process is burdensome to ordering providers and limits the richness and detail of the information communicated to radiologists. This project seeks to add artificial intelligence natural language processing functionality to the CDS software so that it can understand the reason for the exam from a free text field, reducing or eliminating the need for extensive series of check boxes.

As PI of this project, I developed the overall project plan and budget, negotiated it with the corporate partner, supervise the data scientist in extraction and deidentification of the clinical data, wrote the IRB application for the retrospective portion of the project, supervised the IRB for the prospective portion of the project, supervise the testing efforts of Katie Grouse, a neurologist investigator, and serve as the key liaison with the corporate partner.

3.	PI	1 % effort	Mongan (PI)
Radiological Society of North America		06/01/2022	08/20/2022

Covid-19 Imaging Data Extraction and Preparation for the Medical Imaging and Data Resource Center — $ 24074 direct/yr 1 — $ 24074 total

A major limitation in the ability to develop artificial intelligence to address Covid-19 imaging is the availability of large, well-curated, datasets of medical images of patients with Covid-19 and matched negatives. Imaging outside the respiratory system, including brain and heart, will be of increasing importance as the focus of research turns from diagnosis of acute Covid to understanding and management of long Covid. This grant provides support for locating, extracting, deidentifying and transmitting UCSF Covid-19 imaging data to the MIDRC project

As PI I have ultimate responsibility for the project, including managing IRB approval, supervising the data scientist and providing a clinical perspective on which data to obtain and how to interpret it.

---

4. Investigator — 10 % effort — Smith-Bindman (PI)

11/01/2021 — 10/31/2026

SYstems Research on Misuse, Overuse, and Underuse of iNTerventions (SYRMOUNT) — $ 3,005,100 total

I direct that artificial intelligence portions of the project. We are developing artificial intelligence models that can automatically classify the reason that and imaging study was performed (e.g. pain, cancer restaging) and the imaging outcome of the study (e.g. no correlate for symptoms of pain, no change in metastases, no evidence of metastases). These models will enable us to analyze and gather statistics on very large numbers of patients without having to do laborious chart review and coding of each patient.

RESEARCH AWARDS - PAST

1. 11-20 — PI — 0 % effort — Mongan (PI)

UCSF Radiology — 11/01/2011 — 10/31/2012

Development of dual contrast-enhanced dual energy computed tomography abdominal imaging — $ 3242.00 direct/yr 1 — $ 3242.00 total

---

2. 1R21EB013816-01 — Consultant — 0 % effort — Yeh (PI)

NIBIB (PI: Benjamin Yeh) — 04/01/2012 — 03/30/2014

Complementary Injectable Tungsten Contrast for Dual Contrast Dual Energy CT — $ 231,750 direct/yr 1 — $ 413,867 total

The goal of this project is development of a tungsten-based intravenous CT contrast agent with acceptable toxicity profile that is distinguishable from simultaneously administered conventional iodinated contrast at dual energy CT. My role as a consultant is to lead the image processing and analysis aspects of the project.

---

3. 2013246 — Investigator — 1 % effort — Weber (PI)

UCSF Center for Healthcare Value (PI: Ellen Weber) — 07/01/2014 — 06/30/2015

RADS: Reducing CT use with Apex Decision Rules		$ 25,000 direct/yr 1	$ 25,000 total

This project incorporates well-validated clinical decision rules for pulmonary embolism testing into the CT angiogram for PE order to decrease the number of scans ordered without decreasing our ability to detect this high-risk disease.

4.	PI	5 % effort	Mongan (PI)
Enlitic, Inc.		8/29/2016	8/29/2017
Evaluation of Deep Learning Classifiers for Lung Cancer		$ 23,775 direct/yr 1	$ 23,775 total

Deep learning neural network techniques have recently shown great promise in image recognition and analysis. Enlitic, Inc has developed deep learning-based classifiers to predict malignancy of lung lesions on chest CT, but the performance of these classifiers has not been sufficiently characterized. This investigation will use retrospective imaging and pathology data from UCSF to evaluate the performance of deep learning classifiers.

As the PI of this project, I designed the study protocol, wrote the IRB application, identified the data to be extracted, supervised deidentification of the data, and worked with investigators at Enlitic to analyze the results of the classifier output and draft manuscripts.

5. R21DK109433	Co-investigator	10 % effort	Chi (PI)
NIDDK		06/01/2016	05/31/2018
Can Contrast-Enhanced Ultrasound Replace Fluoroscopic Nephrostogram		$ 150,000 direct/yr 1	$ 275,000 total

Fluoroscopic nephrostogram is commonly used after percutaneous nephrolithotomy to assess ureteral patency prior to removal of nephrostomy tubes. While this diagnostic procedure is effective, it involves exposure to ionizing radiation and since it requires a fluoroscopy suite it is expensive and can be difficult to schedule. Ultrasound contrast material can be injected through the nephrostomy tube and its passage into the bladder can be assessed with an ultrasound machine. This investigation compares the efficacy of contrast-enhanced ultrasound to conventional fluoroscopy.

My role includes design of the study, particularly power calculation, developing the contrast ultrasound protocol, performing contrast ultrasound procedures, instructing others in performance of the procedure, statistical analysis of results and drafting manuscripts.

6. R01CA181191	Consultant	1 % effort	Smith-Bindman (PI)
National Cancer Institute		09/12/2014	12/31/2018
CT Dose Collaboratory		$ 904,878 direct/yr 1	$ 4,816,984 total

Computed tomography (CT) is frequently used for medical scanning and can deliver high doses of radiation to patients. Yet because few standards exist for CT examinations, the radiation doses that patients receive during CT vary widely. Routinely, doses are higher than needed for medical diagnoses and high enough to be associated with increased cancer risk. The proposed project is a multisite collaboration studying improved standards for conducting CT, including the radiation doses used, and developing strategies to apply (implement) and spread (disseminate) these standards in different clinics and hospitals. The project will use a mixture of methods including a randomized controlled trial, observational data, and key informant interviews. The work strives to improve CT safety and reduce cancers associated with radiation from CT by lowering the doses that patients receive.

As an expert on abdominopelvic CT, CT dose reduction and applied CT physics, I advise the investigators on presentation of CT dose data and study participants on CT protocols and dose reduction techniques.

7.		Co-PI	25 % effort	Mongan/Taylor (PI)
	General Electric		10/01/2016	06/30/2018
	Screening for Dangerous Findings on Chest Radiographs		$ 2,969,452 direct/yr 1	$ 2,969,452 total

This is a sub-award of 10-year $13 million master agreement with General Electric. This sub-award focuses on using deep learning neural network techniques to identify two common and potentially acutely dangerous conditions seen on chest radiographs: misplaced feeding tubes and pneumothorax. The goal is to create a classifier that would improve patient care in the acute setting by analyzing every radiograph and flagging those that are likely to have one of these findings for priority review by a radiologist.

As Co-PI of this study, my role is to define the clinical question, identify and prepare input data, design the compute infrastructure needed to perform the research, define the approach to creating the classifier, construct a rigorous evaluation of performance, perform statistical analysis of the results, and draft manuscripts.

8.	OOS030101	Consultant	1 % effort	Smith-Bindman (PI)
	PCORI		08/01/2015	7/31/2020
	Pragmatic Trial of More versus Less Intensive Strategies for Active Surveillance of Patients with Small Pulmonary Nodules		$ 224,885 direct/yr 1	$ 893,707 total

Lung nodules are a common finding on CTs that include the lungs. The majority of small nodules are benign, but a small percentage develop malignancies. This investigation seeks to determine the effectiveness of different follow-up strategies in maximizing quality of care.

As an informatics expert, I advise the PI and her team on development of software for decision support, particularly user interface design, and data collection.

9.		PI	1 % effort	Mongan (PI)
	General Electric		01/01/2020	07/31/2021
	Medical Evaluation of the Critical Care Suite on the Optima XR240amx mobile X-Ray system		$ 31,755.48 direct/yr 1	$ 31,755.48 total

I led the development of an AI-based algorithm to detect pneumothorax, published in PLoS: Medicine. GE has licensed this work, implemented it on a portable X-ray machine and obtained FDA clearance for the device. This grant supports integration of the AI-enabled X-ray machine into the clinical PACS to enable worklist prioritization based on AI results and evaluation of the experience of using the AI in a clinical environment.

Negotiate contract, direct informatics integration, analyze results and outcomes, draft whitepaper.

PEER REVIEWED PUBLICATIONS

1. **Mongan J**. Interactive essential dynamics. J Comput Aided Mol Des. 2004 Jun; 18(6):433-6. PMID: 15663003
2. Hamelberg D, **Mongan J**, McCammon JA. Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. J Chem Phys. 2004 Jun 22; 120(24):11919-29. PMID: 15268227
3. **Mongan J**, Case DA, McCammon JA. Constant pH molecular dynamics in generalized Born implicit solvent. J Comput Chem. 2004 Dec; 25(16):2038-48. PMID: 15481090
4. **Mongan J**, Case DA. Biomolecular simulations at constant pH. Curr Opin Struct Biol. 2005 Apr; 15(2):157-63. PMID: 15837173
5. Swanson JM, **Mongan J**, McCammon JA. Limitations of atom-centered dielectric functions in implicit solvent models. J Phys Chem B. 2005 Aug 11; 109(31):14769-72. PMID: 16852866
6. Puerta DT, **Mongan J**, Tran BL, McCammon JA, Cohen SM. Potent, selective pyrone-based inhibitors of stromelysin-1. J Am Chem Soc. 2005 Oct 19; 127(41):14148-9. PMID: 16218585
7. Lewis JA*, **Mongan J***, McCammon JA, Cohen SM. Evaluation and binding-mode prediction of thiopyrone-based inhibitors of anthrax lethal factor. ChemMedChem. 2006 Jul; 1(7):694-7. PMID: 16902919 (* Co-primary authors)
8. **Mongan J**, Simmerling C, McCammon JA, Case DA, Onufriev A. Generalized Born model with a simple, robust molecular volume correction. J Chem Theory Comput. 2007 Jan 1; 3(1):156-169. PMID: 21072141.
9. **Mongan J**, Svrcek-Seiler WA, Onufriev A. Analysis of integral expressions for effective Born radii. J Chem Phys. 2007 Nov 14; 127(18):185101. PMID: 18020664
10. **Mongan J**, Rathnayake S, Fu Y, Wang R, Jones EF, Gao DW, Yeh BM. In vivo differentiation of complementary contrast media at dual-energy CT. Radiology. 2012 Oct; 265(1):267-72. PMID: 22778447.
11. Coakley FV, Hanley-Knutson K, **Mongan J**, Barajas R, Bucknor M, Qayyum A. Pancreatic imaging mimics: part 1, imaging mimics of pancreatic adenocarcinoma. AJR Am J Roentgenol. 2012 Aug; 199(2):301-8. PMID: 22826390
12. **Mongan J**, Rathnayake S, Fu Y, Gao DW, Yeh BM. Extravasated contrast material in penetrating abdominopelvic trauma: dual-contrast dual-energy CT for improved diagnosis--preliminary results in an animal model. Radiology. 2013 Sep; 268(3):738-42. PMID: 23687174.

13. Yu JP, Kansagra AP, **Mongan J**. The radiologist's workflow environment: evaluation of disruptors and potential implications. J Am Coll Radiol. 2014 Jun; 11(6):589-93. PMID: 24775910

14. **Mongan J**, Kline J, Smith-Bindman R. Age and sex-dependent trends in pulmonary embolism testing and derivation of a clinical decision rule for young patients. Emerg Med J. 2015 Mar 9. PMID: 25755270

15. *Rathnayake S, **Mongan J**, Torres AS, Colborn R, Gao DW, Yeh BM, Fu Y. Rathnayake S, Mongan J, Torres AS, Colborn R, Gao DW, Yeh BM, Fu Y. In vivo comparison of tantalum, tungsten, and bismuth enteric contrast agents to complement intravenous iodine for double-contrast dual-energy CT of the bowel. Contrast Media Mol Imaging. 2016 Feb 18. PMID: 26892945 (*First author was a trainee who I mentored on this project.)

16. *Li Y, **Mongan J**, Behr SC, Sud S, Coakley FV, Simko J, Westphalen AC. Beyond Prostate Adenocarcinoma: Expanding the Differential Diagnosis in Prostate Pathologic Conditions. Radiographics. 2016 Jul-Aug; 36(4):1055-75. PMID: 27315446 (*First author was a trainee who I mentored on this project.)

17. **Mongan J,** Sebro R. Definition of Confidence Interval. Radiographics. 2016 Sep-Oct; 36(5):1602. PMID: 27618334

18. Usawachintachit M, Tzou DT, **Mongan J**, Taguchi K, Weinstein S, Chi T. Feasibility of Retrograde Ureteral Contrast Injection to Guide Ultrasonographic Percutaneous Renal Access in the Nondilated Collecting System. J Endourol. 2017 Feb; 31(2):129-134. PMID: 27809568. PMCID: PMC5312625

19. Chi T, Usawachintachit M, **Mongan J**, Kohi MP, Taylor A, Jha P, Chang HC, Stoller M, Goldstein R, Weinstein S. Feasibility of Antegrade Contrast-enhanced US Nephrostograms to Evaluate Ureteral Patency. Radiology. 2017 Apr; 283(1):273-279. PMID: 28234551. PMCID: PMC5375626

20. Rajkomar A, Lingam S, Taylor AG, Blum M, **Mongan J**. High-Throughput Classification of Radiographs Using Deep Convolutional Neural Networks. J Digit Imaging. 2017 Feb; 30(1):95-101. PMID: 27730417. PMCID: PMC5267603

21. Phelps A, Callen AL, Marcovici P, Naeger DM, **Mongan J**, Webb EM. Can Radiologists Learn From Airport Baggage Screening?: A Survey About Using Fictional Patients for Quality Assurance. Acad Radiol. 2017 Nov 06. PMID: 29122472

22. Usawachintachit M, Tzou DT, **Mongan J**, Weinstein S, Chi T. Antegrade ultrasound contrast injection facilitates accurate nephrostomy tube positioning during percutaneous nephrolithotomy. Int J Urol. 2017 Mar; 24(3):239-240. PMID: 27862356

23. Marcus SG, Candia S, Kohli MD, **Mongan J**, Zagoria RJ, Behr SC, Sun D, Westphalen AC. Association between misty mesentery with baseline or new diagnosis of cancer: a matched cohort study. Clin Imaging. 2017 Dec 15; 50:57-61. PMID: 29276962

24. Chi T, Usawachintachit M, Weinstein S, Kohi MP, Taylor A, Tzou DT, Chang HC, Stoller M, **Mongan J**. Contrast Enhanced Ultrasound as a Radiation Free Alternative to Fluoroscopic Nephrostogram for Evaluating Ureteral Patency. J Urol. 2017 Jul 23. PMID: 28743528

25. Tzou DT, Weinstein S, Usawachintachit M, **Mongan J**, Greene KL, Chi T. Contrast Enhanced Ultrasound Detects Recurrent Renal Cell Carcinoma in the Setting of Chronic Renal Insufficiency. Clin Genitourin Cancer. 2017 Aug; 15(4):e735-e737. PMID: 28131753. PMCID: PMC5540333

26. Kovacs MD, Mesterhazy J, Avrin D, Urbania T, **Mongan J**. Correlate: A PACS- and EHR-integrated Tool Leveraging Natural Language Processing to Provide Automated Clinical Follow-up. Radiographics. 2017 Sep-Oct; 37(5):1451-1460. PMID: 28898194

27. **Mongan J**, Avrin D. Impact of PACS-EMR Integration on Radiologist Usage of the EMR. J Digit Imaging. 2018 Apr 25. PMID: 29696473

28. Taylor AG, Mielke C, **Mongan J**. Automated detection of moderate and large pneumothorax on frontal chest X-rays using deep convolutional neural networks: A retrospective study. PLoS Med. 2018 Nov; 15(11):e1002697. PMID: 30457991. PMCID: PMC6245672

29. Fahimi J, Kanzaria HK, **Mongan J**, Kahn KL, Wang RC. Potential Effect of the Protecting Access to Medicare Act on Use of Advanced Diagnostic Imaging in the Emergency Department: An Analysis of the National Hospital Ambulatory Care Survey. Radiology. 2019 Jan 29; 181650. PMID: 30694161

30. Hentel KD, Menard A, Mongan J, Durack JC, Raja AS, Khorasani R. Mandated Imaging Appropriate Use Criteria. Ann Intern Med. 2019 11 05; 171(9):682-683. PMID: 31683282

31. Hentel KD, Menard A, **Mongan J**, Durack JC, Johnson PT, Raja AS, Khorasani R. What Physicians and Health Organizations Should Know About Mandated Imaging Appropriate Use Criteria. Ann Intern Med. 2019 Jun 11. PMID: 31181572

32. Kallianos K, **Mongan J**, Antani S, Henry T, Taylor A, Abuya J, Kohli M. How far have we come? Artificial intelligence for chest radiograph interpretation. Clin Radiol. 2019 Jan 28. PMID: 30704666

33. **Mongan J**, Kohli M. Artificial Intelligence and Human Life: Five Lessons for Radiology from the 737 MAX Disasters. Radiol Artif Intell. 2020 Mar; 2(2):e190111. PMID: 33937819. PMCID: PMC8017379

34. **Mongan J**, Moy L, Kahn CE. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): A Guide for Authors and Reviewers. Radiol Artif Intell. 2020 Mar; 2(2):e200029. PMID: 33937821. PMCID: PMC8017414

35. Flanders AE, Prevedello LM, Shih G, Halabi SS, Kalpathy-Cramer J, Ball R, **Mongan JT**, Stein A, Kitamura FC, Lungren MP, Choudhary G, Cala L, Coelho L, Mogensen M, Morón F, Miller E, Ikuta I, Zohrabian V, McDonnell O, Lincoln C, Shah L, Joyner D, Agarwal A, Lee RK, Nath J, RSNA-ASNR 2019 Brain Hemorrhage CT Annotators . Construction of a Machine Learning Dataset through Collaboration: The RSNA 2019 Brain CT Hemorrhage Challenge. Radiol Artif Intell. 2020 May; 2(3):e190211. PMID: 33937827. PMCID: PMC8082297

36. Filice RW, **Mongan J**, Kohli MD. Evaluating Artificial Intelligence Systems to Guide Purchasing Decisions. J Am Coll Radiol. 2020 Nov; 17(11):1405-1409. PMID: 33035503

37. Houshyar R, Tran-Harding K, Glavis-Bloom J, Nguyentat M, **Mongan J**, Chahine C, Loehfelm TW, Kohli MD, Zaragoza EJ, Murphy PM, Kampalath R. Effect of shelter-in-place on emergency department radiology volumes during the COVID-19 pandemic. Emerg Radiol. 2020 Dec; 27(6):781-784. PMID: 32504280. PMCID: PMC7273127

38. Tsai EB, Simpson S, Lungren M, Hershman M, Roshkovan L, Colak E, Erickson BJ, Shih G, Stein A, Kalpathy-Cramer J, Shen J, Hafez M, John S, Rajiah P, Pogatchnik BP, **Mongan J**, Altinmakas E, Ranschaert ER, Kitamura FC, Topff L, Moy L, Kanne JP, Wu

CC. The RSNA International COVID-19 Open Annotated Radiology Database (RICORD). Radiology. 2021 Jan 05; 203957. PMID: 33399506. PMCID: PMC7993245

39. Colak E, Kitamura FC, Hobbs SB, Wu CC, Lungren MP, Prevedello LM, Kalpathy-Cramer J, Ball RL, Shih G, Stein A, Halabi SS, Altinmakas E, Law M, Kumar P, Manzalawi KA, Nelson Rubio DC, Sechrist JW, Germaine P, Lopez EC, Amerio T, Gupta P, Jain M, Kay FU, Lin CT, Sen S, Revels JW, Brussaard CC, **Mongan J**, RSNA-STR Annotators and Dataset Curation Contributors . The RSNA Pulmonary Embolism CT Dataset. Radiol Artif Intell. 2021 Mar; 3(2):e200254. PMID: 33937862. PMCID: PMC8043364

40. Eng D, Chute C, Khandwala N, Rajpurkar P, Long J, Shleifer S, Khalaf MH, Sandhu AT, Rodriguez F, Maron DJ, Seyyedi S, Marin D, Golub I, Budoff M, Kitamura F, Takahashi MS, Filice RW, Shah R, **Mongan J**, Kallianos K, Langlotz CP, Lungren MP, Ng AY, Patel BN. Automated coronary calcium scoring using deep learning with multicenter external validation. NPJ Digit Med. 2021 Jun 01; 4(1):88. PMID: 34075194. PMCID: PMC8169744

41. **Mongan J**, Kalpathy-Cramer J, Flanders A, George Linguraru M. RSNA-MICCAI Panel Discussion: Machine Learning for Radiology from Challenges to Clinical Applications. Radiol Artif Intell. 2021 Sep; 3(5):e210118. PMID: 34617032. PMCID: PMC8489458

42. Webb EM, **Mongan J**. Gastrointestinal Stromal Tumors: Radiomics may Increase the Role of Imaging in Malignant Risk Assessment. Acad Radiol. 2022 Mar 02. PMID: 35248459

43. **Mongan J**, Kohli M, Houshyar R, Chang P, Glavis-Bloom J, Taylor A. Automated Detection of IVC Filters with Deep Convolutional Neural Networks. Submitted.

BOOKS AND CHAPTERS

1. Programming Interviews Exposed, by **John Mongan** and Noah Suojanen; Wiley, 2000. Second edition by **John Mongan**, Noah Suojanen and Eric Giguere; Wiley, 2007. Third edition by **John Mongan**, Eric Giguere and Noah Kindler; Wiley, 2012. Fourth edition by **John Mongan**, Noah Kindler and Eric Giguere; Wiley 2018. (Over 100,000 copies sold). Republished in China (Simplified Chinese), South Korea (Korean), India (English), Russia (Russian), and Poland (Polish).

2. **Mongan J.** Contributor of 10 cases (approx. 5% of total content) to Gastrointestinal Imaging Cases, edited by Angela D. Levy, Koenraad Mortele and Benjamin M. Yeh; Oxford University Press; 2013.

SIGNIFICANT PUBLICATIONS

1. Rajkomar A, Lingam S, Taylor AG, Blum M, **Mongan J**. High-Throughput Classification of Radiographs Using Deep Convolutional Neural Networks. J Digit Imaging. 2016 Oct 11. PMID: 27730417

   This proof-of-concept work was one of the early applications of the current generation of convolutional neural network artificial intelligence techniques to radiological images, demonstrating that very high levels of accuracy can be obtained on simple but non-trivial computer vision problems in medical imaging. I framed the scientific problem, obtained the data sets, set scientific direction, determined analysis strategy, outlined strategy for manuscript, and selected journal for submission.

2. **Mongan J**, Avrin D. Impact of PACS-EMR Integration on Radiologist Usage of the EMR. J Digit Imaging. 2018 Apr 25. PMID: 29696473

   This investigation quantified the impact of the Radiant integration between Epic/Apex and the PACS software used by radiologists to view imaging on radiologist's use of Apex; it employed security audit logs as a novel data source to understand usage patterns of medical software. I took the lead role in all aspects of this project, including framing the scientific problem; performing literature search; designing the investigation; writing the IRB; collecting, processing and analyzing the data; and drafting the manuscript.

3. Taylor AG, Mielke C, **Mongan J**. Automated detection of moderate and large pneumothorax on frontal chest X-rays using deep convolutional neural networks: A retrospective study. PLoS Med. 2018 Nov; 15(11):e1002697. PMID: 30457991. PMCID: PMC6245672

   This was an artificial intelligence development project that aimed to create a pneumothorax detection algorithm with sufficiently high performance to be clinically useful. The work was published in a special artificial intelligence issue of PLoS: Medicine; the intellectual property developed was licensed to GE and has been incorporated into a portable X-ray device. I framed the scientific problem, wrote the software used for annotation of over 10,000 radiographs, organized and analyzed the annotations, directed the data science work of algorithm development, determined the analysis strategy, outlined the strategy for the manuscript, revised and refined the manuscript, selected journal for submission and revised the manuscript based on reviewer and editor comments.

4. **Mongan J**, Kohli M. Artificial Intelligence and Human Life: Five Lessons for Radiology from the 737 Max Disasters. Radiology: Artificial Intelligence. 2020 Mar 18.

   This article looks toward the translational implementation of artificial intelligence in clinical medicine and seeks to avoid potential morbidity and mortality from missteps by drawing lessons from recent aviation disasters. I conceived the concept, performed background literature search, drafted four of the five lessons, revised and refined the manuscript and submitted the manuscript.

5. **Mongan J**, Moy L, Kahn CE. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): A Guide for Authors and Reviewers. Radiology: Artificial Intelligence. 2020 Mar 25

   This article describes a checklist developed to improve the quality of publications on artificial intelligence in medical imaging by identifying key pieces of information that must be included and indicators of quality of research. It has been adopted by the journal Radiology: Artificial Intelligence as a mandatory part of the submission process for all original research articles. I identified the need for the checklist, drafted the first outline, worked with coauthors to develop all the points of the checklist, and edited and revised the final manuscript.

PATENTS ISSUED OR PENDING

1. 6,189,116: Complete, Randomly Ordered Traversal of Cyclic Directed Graphs. **John T. Mongan** and Dorothy M. Cribbs.
2. 6,304,982: Network Distributed Automated Testing System. **John T. Mongan**, Dorothy M. Cribbs and John R. DeAguiar.
3. 6,378,088: Monte Carlo Automated Test Generator. **John T. Mongan.**

CONFERENCE ABSTRACTS

1. * "Constant pH Molecular Dynamics in Generalized Born Implicit Solvent" 229th American Chemical Society National Meeting in San Diego, California. **John Mongan**, David A. Case. (Delivered formal oral presentation)
2. "Methods for Radiation Dose Reduction in Abdominopelvic CT Imaging." Annual Meeting of the Radiological Society of North America Chicago, Illinois. **John Mongan**, Rizwan Aslam, Fergus Coakley, Robert Gould, John Shepherd, Benjamin Yeh. (poster, selected for CME credit)
3. * "Trauma Imaging with Color Contrast for Color CT: In Vivo Use of Complementary Contrast Materials at Dual-energy Computed Tomography" **John Mongan**, Samira Rathnayake, Yanjun Fu, Benjamin Yeh. Annual Meeting of the Radiological Society of North America Chicago, Illinois. (Delivered formal oral presentation)
4. "Improving Efficiency of Pulmonary Embolism Testing in Young Female Patients" **John Mongan**, Jeffrey Kline, Rebecca Smith-Bindman. Annual Meeting of the Radiological Society of North America in Chicago, Illinois. (poster, alternate for oral presentation)
5. * "Dual Contrast Dual-Energy CT in Imaging of Penetrating Abdominal Trauma" **John Mongan**, Samira Rathnayake, Yanjun Fu, Benjamin Yeh. Contrast2013 Contrast Media Research Symposium Beijing, China. (Delivered formal oral presentation)
6. "Beyond Prostate Cancer: Expanding the Differential Diagnosis in Prostate Pathology" Yi Li**, **John Mongan**, Spencer Behr, Seema Sud, Fergus Coakley, Jeff Simko, Antonio Westphalen. Annual Meeting of the Radiological Society of North America Chicago, Illinois. (poster; selected for Radiographics publication and CME credit. First author was a trainee mentored and supervised by me)
7. "The Radiologist's Workflow Environment: Evaluation of Disruptors and Potential Implications" JP Yu, Akash Kansagra, **John Mongan**. Annual Meeting of the Radiological Society of North America Chicago, Illinois. (Formal oral presentation)
8. "Computerized Provider Order Entry (CPOE) as a Cause of Errors in Imaging Requests: What a Difference a Space Makes" **John Mongan**, Aaron Neinstein, Christopher Jovais, Spencer Behr. Annual Meeting of the Radiological Society of North America Chicago, Illinois. (poster; selected for CME credit)
9. ** "Correlate: A PACS and EMR-integrated Tool which Leverages Natural Language Processing (NLP)to Provide Automated Clinical Follow-up for the Radiologist" Mark D. Kovacs, Joseph Mesterhazy, David E. Avrin, Thomas H. Urbania, **John Mongan** Annual Meeting of the Radiological Society of North America Chicago, Illinois. (poster; selected for CME credit. First author was a trainee mentored and supervised by me)
10. "Improving Automated Clinical Follow-up Through Optimization of Natural Language Processing (NLP) Methods" Mark D. Kovacs, Joseph Mesterhazy, David E. Avrin, Thomas H. Urbania, **John Mongan.** Annual Meeting of the Society for Imaging Informatics in Medicine, Portland, Oregon
11. "Context integration of EMR (electronic medical record) with PACS increases radiologist use of EMR" **John Mongan**, David Avrin. Annual Meeting of the Society for Imaging Informatics in Medicine, Portland, Oregon

12. "Decision Support Tool in an EMR Improves Ordering" Ellen Weber, Ralph Wang, **John Mongan**, et al. Royal College of Emergency Medicine Annual Scientific Conference, Bornemouth, United Kingdom.
13. "Feasibility of antegrade contrast-enhanced ultrasound nephrostograms" Manint Usawachintachit, **John Mongan**, Stefanie Weinstein, Thomas Chi. World Congress of Endourology. Cape Town, South Africa.
14. ** "Targeted QA: Creating a PACS based Teaching File using Pareto Analysis of Trainee Discrepancies" Hriday Shah, **John Mongan**, Eric Ehman, Javier Villanueva-Meyer, Soonmee Cha, Jason Talbott. Annual Meeting of the Radiological Society of North America Chicago, Illinois. (Formal oral presentation)
15. ** "A Shiny New World: Creating Your Own Radiology Decision Support Webapps Using R" Jennifer Wan, **John Mongan**, David Incerti, Jesse Courtier. Annual Meeting of the Radiological Society of North America Chicago, Illinois. (poster; selected for CME credit)
16. ** "Clinical Decision Support and Appropriate Use Criteria: Complying with PAMA while Improving Usefulness of Imaging" Matthew Barkovich, Marc Kohli, William Dillon, Rebecca Smith-Bindman, **John Mongan**. Annual Meeting of the Radiological Society of North America Chicago, Illinois.
17. ** "Convolutional Neural Networks: Fundamental Theory and Guided Implementation for the Radiologist Using Google Collaboratory" Alex Chan, Spencer Behr, **John Mongan.** Annual Meeting of the Radiologicial Society of North America Chicago, Illinois. (poster; selected for Radiographics publication. First author was a trainee mentored and supervised by me.)
18. "Synthetic Post-contrast Breast MRI for Imaging Malignant Lesions Using Deep Learning" Maggie Chung, Evan Calabrese, Kimberly Ray, Tatiana Kelil, Genevieve Woodard, Nola Hylton, **John Mongan**, Bonnie Joe, Amie Lee. Annual Meeting of the Radiologicial Society of North America Chicago, Illinois. (poster)
19.
    * Formal oral presentation by me
    ** First author was a trainee mentored and supervised by me

ACADEMIC LEADERSHIP

My primary academic leadership role is serving **Associate Chair, Translational Informatics** for Radiology and Biomedical Imaging. In conjunction with the Associate Chair, Clinical Informatics, I **lead the Imaging IT group, consisting of approximately 25 FTE**, including three full-time IT directors who jointly report directly to us.

I have created a **comprehensive infrastructure for artificial intelligence** work in radiology. Building on earlier success in the creation of the Automated Image Retrieval (AIR) platform, which provides web-based self-service access to deidentification and extraction of image data files from the clinical PACS archive, I directed the **expansion in functionality of AIR** to include more flexible deidentification profiles and the extraction and deidentification of radiology reports. Two additional AIR instances are **deployed at the SF VAMC and ZSFG**, providing image data file access across all medical centers staffed by UCSF faculty. I negotiated a **co-funding agreement with UCD Radiology** and we have installed an instance at UCD as well. Recognizing a common need for efficient annotation of images, I negotiated a reduced-price contract for the **MD.ai cloud-based annotation** platform which has found wide application within the department for AI projects as well as traditional imaging investigations

requiring multi-center reads. I led an **overhaul of the hardware infrastructure** managed by Radiology Scientific Computing Services, including **transitioning storage** from unreliable consumer-grade equipment to a NetApp storage device, **consolidating data back-up** at Parnassus to increase efficiency and **dramatically increasing GPU compute** capacity, including purchase and deployment of a 16-GPU NVIDIA DGX-2 server. In parallel with hardware infrastructure, I have directed a modernization of our research software infrastructure including **GPU-enabling our compute scheduling** system and containerizing our AI and GPU frameworks. Looking toward clinical deployment of AI, I led the selection, purchase and implementation of a **DICOM router system**, which allows rules to be created for **automatically forwarding selected imaging data to AI** processing pipelines as soon as it is scanned.

Recognizing the importance of sharing experience and leveraging scale for value across UC, **I continue to organize regular UC Radiology Informatics** meetings with my Radiology informatics counterparts at the other four UC medical campuses. We have been meeting regularly for almost 6 years. At these meetings, we discuss current and potential shared UC-wide projects (including enterprise imaging, clinical decision support, AIR and UC-wide PACS) and share lessons learned and successful strategies for addressing the similar challenges that we face at each of our campuses.

In conjunction with Marc Kohli, Director of Clinical Informatics **I continue to supervise the IT and informatics infrastructure of the Department of Radiology and Biomedical Imaging**. This includes the newly updated clinical PACS systems (Visage Viewer, Nuance Workflow Orchestration and Communicator, DynaCAD for prostate and breast MR, MIM for general nuclear medicine), the Epic Radiant RIS system, Powerscribe 360 dictation, Enterprise Imaging (eUnity "webpacs" viewer software and UC-wide vendor-neutral archive), two Radiology data centers (including multi petabyte storage systems, virtualized compute servers and backup systems), network and firewall equipment, a research computing cluster, and a variety of educational software systems (UCSF teaching file server, "wet-read" preliminary report system, and mPower report search and analytics tool). My responsibilities include strategic planning, personnel decisions, budgeting, and prioritization. I have weekly meetings with the Imaging IT group (25 FTE), twice monthly meetings with the Scientific Computing Services (3, soon to be 4 FTE) group and direct interaction by phone and email with these groups multiple times per day.

OTHER CREATIVE ACTIVITIES

1. **SOFTWARE**

2. AMBER 8. D.A. Case, T.A. Darden, T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, K.M. Merz, B. Wang, D.A. Pearlman, M. Crowley, S. Brozell, V. Tsui, H. Gohlke, **J. Mongan**, V. Hornak, G. Cui, P. Beroza, C. Schafmeister, J.W. Caldwell, W.S. Ross, and P.A. Kollman. 2004. http://ambermd.org

3. Interactive Essential Dynamics 2.0. **John Mongan**. July 2004. http://mccammon.ucsd.edu/ied

4. AMBER 9. D.A. Case, T.A. Darden, T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, K.M. Merz, D.A. Pearlman, M. Crowley, R.C. Walker, W. Zhang, B. Wang, S. Hayik, A. Roitberg, G. Seabra, K.F. Wong, F. Paesani, X. Wu, S. Brozell, V. Tsui, H. Gohlke, L. Yang, C. Tan, **J. Mongan**, V. Hornak, G. Cui, P. Beroza, D.H. Mathews, C. Schafmeister, W.S. Ross, and P.A. Kollman. 2006. http://ambermd.org

5. AMBER 10. D.A. Case, T.A. Darden, T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, M. Crowley, R.C. Walker, W. Zhang, K.M. Merz, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossváry, K.F. Wong, F. Paesani, J. Vanicek, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, L. Yang, C. Tan, **J. Mongan**, V. Hornak, G. Cui, D.H. Mathews, M.G. Seetin, C. Sagui, V. Babin, and P.A. Kollman. 2008. http://ambermd.org
6. **Ultrasound-guided biopsy curriculum**
   Recorded one hour lecture illustrating step by step procedure for preparing patient and performing successful ultrasound guided biopsies. This is required viewing for each resident prior to starting the Mt. Zion Abdominal Imaging rotation.

EXHIBIT R-2

Clinical Radiology 76 (2021) 317–324

Contents lists available at ScienceDirect

Clinical Radiology

journal homepage: www.clinicalradiologyonline.net

Review

Artificial intelligence in radiology: relevance of collaborative work between radiologists and engineers for building a multidisciplinary team

T. Martín-Noguerol [a,*], F. Paulano-Godino [b], R. López-Ortega [b], J.M. Górriz [c], R.F. Riascos [d], A. Luna [a]

[a] *MRI Unit, Radiology Department. HT Medica, Jaén, Spain*
[b] *Engineering Department, HT Medica, Jaén, Spain*
[c] *Department of Signal Theory, Telematics and Communications, University of Granada, Granada, Spain*
[d] *Department of Neuroradiology, The University of Texas Health Science Center at Houston, McGovern Medical School, Houston, TX, USA*

ARTICLE INFORMATION

Article history:
Received 11 September 2020
Accepted 20 November 2020

The use of artificial intelligence (AI) algorithms in the field of radiology is becoming more common. Several studies have demonstrated the potential utility of machine learning (ML) and deep learning (DL) techniques as aids for radiologists to solve specific radiological challenges. The decision-making process, the establishment of specific clinical or radiological targets, the profile of the different professionals involved in the development of AI solutions, and the relation with partnerships and stakeholders are only some of the main issues that have to be faced and solved prior to starting the development of radiological AI solutions. Among all the players in this multidisciplinary team, the communication between radiologists and data scientists is essential for a successful collaborative work. There are specific skills that are inherent to radiological and medical training that are critical for identifying anatomical or clinical targets as well as for segmenting or labelling lesions. These skills would then have to be transferred, explained, and taught to the data science experts to facilitate their comprehension and integration into ML or DL algorithms. On the other hand, there is a wide range of complex software packages, deep neural-network architectures, and data transfer processes for which radiologists need the expertise of software engineers and data scientists in order to select the optimal manner to analyse and post-process this amount of data. This paper offers a summary of the top five challenges faced by radiologists and data scientists including tips and tricks to build a successful AI team.

* Guarantor and correspondent: T. Martín-Noguerol, MRI Section, Radiology Department, HT Medica, Carmelo Torres 2, 23007 Jaén, Spain. Tel.: +34 953275601; fax: +34 953275609.
E-mail address: t.martin.f@htime.org (T. Martín-Noguerol).

https://doi.org/10.1016/j.crad.2020.11.113
0009-9260/

Introduction

Currently, the integration of Artificial intelligence (AI) into radiology is a reality rather than a future promise.[1] There is a growing number of clinical applications based on AI algorithms, that enable radiologists to improve their diagnostic accuracy and have a direct impact on patient healthcare.[2–4] Before they can be used in common radiological practice, the development and validation processes for these AI applications require strong collaboration between radiologists, software engineers, and data science experts. Moreover, the investigation of new and more advanced machine learning (ML), and especially deep learning (DL), algorithms benefits from multidisciplinary teams of radiologists, engineers, and data scientists.[5,6] This demand for conjoined work arises from the radiologists' inevitable lack of technical knowledge of ML and DL approaches, as their academic and clinical training is mostly focused on medicine and physiopathology.[7] In a similar manner, engineers and data science experts have deep gaps in their education about the pathophysiological meaning, radiological and anatomical concepts, and biological variables that may influence the results of their proposed algorithms.[8] On one hand, most practicing radiologists cannot meet the requirements to understand the "black box", which underlies most AI algorithms.[9,10] This fact limits the radiologist's role in the development of radiological AI algorithms, and in the very important process of explaining AI decisions to other physicians, and particularly patients, which has been called explainable AI.[11,12] On the other hand, engineers and data scientists involved in the development of radiological AI-based solutions have similar issues in their daily work: there are many anatomical and biological concepts, radiological features of normal and pathological tissues, and different therapeutic options that hinder their understanding of the target for the algorithm they are developing, and these different factors can influence the final result.[13] For these reasons, collaborative work between this multidisciplinary team is essential. Understanding the specific requirements of the team member is required for the development of successful AI applications. This step will be essential for faster introduction of clinically significant AI-based tools in the field of radiology.[14] Nevertheless, there are other fields whereas AI is showing promising results, such as administrative processes with high economic impact, protocolisation, or image acquisition. In this review, we will focus on the imaging interpretation and natural language applications of AI in radiology. To this aim, the specific requirements of each of the parties involved will be discussed from their respective viewpoints, and a perspective with tips and tricks for building a successful multidisciplinary AI team will be detailed.

What radiologists need from engineers and data scientists

Data visualisation and interpretation

When using AI-based solutions, engineers must be able to explain their results to radiologists.[15,16] In medicine, where the decisions made are critical and directly affect patients and the health system, it is of paramount importance that radiologists understand the decision process of AI algorithms. DL is becoming used more widely in radiology,[17] but its lack of interpretability and transparency is a problem. This is especially problematic for end-users who must trust the results from deep neural networks to make critical decisions.[18] This is generally known as the "black box" of AI.[19,20] Data visualisation can help radiologists and engineers to understand and improve DL models. As an example, deconvolutional networks enable projection from the model's learned feature space back to the pixel space.[21] Interactive techniques have also been incorporated to these visualisation techniques, such as deep inside convolutional neural networks (CNN)[22] or computer-aided systems (CAD) based on visual support.[15,23–25] Therefore, visual analytics must be an integral part of any DL model in order to enhance the radiologists' understanding as well as improve communication between specialists and, particularly, patients. This problem is resolved, as mentioned before, by explainable AI, and it is expected to grow in a manner parallel to that of the use of AI in radiology.[26]

Data preprocessing and feature engineering

ML requires data to learn and build patterns that are used in regression or classification tasks.[27,28] Data preprocessing is a key step in any data science project. Data scientists spend around 80% of their time in data preparation, a time that will exponentially increase if it depends on radiologists' work.[29] Using a considerable quantity of radiological data to train a CNN is computationally expensive. For medical applications, these architectures need to handle high-resolution three-dimensional (3D) images, thus making the training process much more costly.[30]

With the increased interest in data-driven algorithms for radiology, it is of paramount importance to be able to use and reuse this increasing amount of information.[31] This introduces a huge challenge in the selection, curation, de-identification, and storage of radiology data.[32] Hence, it is essential that the quality of the data is sufficient to move forward to AI applications (Fig 1).[33]

Data concerning health involves personal data, such as physical health status. This type of data falls into a special category called "sensitive", meaning that these data should have special protection.[34,35] A Digital Imaging and Communication On Medicine (DICOM) file not only contains a viewable image, but also contains a large variety of data elements.[36] These meta-data elements include identifiable information about the patient, the study, and the institution. Because of this, special care must be taken when sharing and using these data. Numerous tools have been built to perform the task of DICOM data de-identification that can be used for this purpose[37]; however, each tool produces its own de-identification process and outcomes. Therefore, a customisable approach, using a programming language such as Python, is often desirable.

To apply ML algorithms to biomedical image data, a preprocessing step is often required. This involves processes

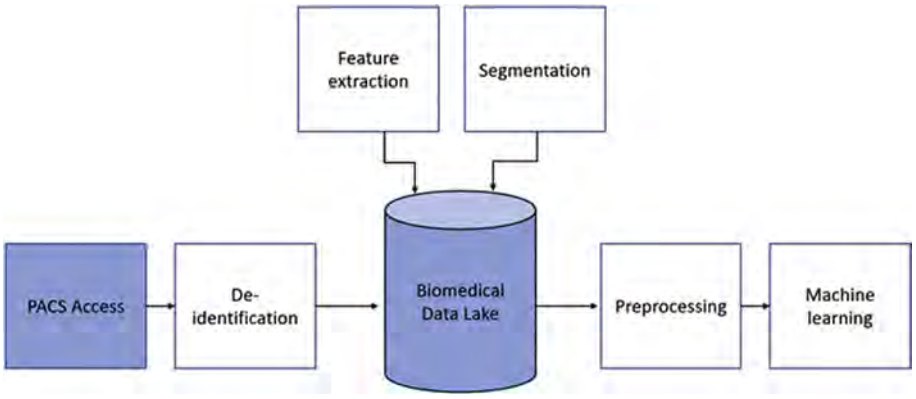


Figure 1 High-level schematic diagram of the data pipeline for radiology images.

such as performing bias field correction in magnetic resonance imaging (MRI) images and image normalisation and resampling.[38] Although with modern DL algorithms important features can be automatically learned, enhancement of these data is still required.[39] In addition, it is worth mentioning the importance of obtaining biomarkers as a previous step before applying ML algorithms. These biomarkers can be processed by reducing their dimensionality, that is, by reducing the number of biomarkers obtained while maintaining the same information they provide.

Thus, high-quality data are essential for training an ML algorithm.[40] In this regard, feature-engineering processes aim to improve data quality, and in turn, enhance the performance of the trained models for radiologists, being capable of properly understanding the resulting outcomes. In contrast, DL models allow systems to use raw data as an input, thereby enabling them to automatically discover highly discriminative features in the given training dataset, which is the basis for radiomics and texture analysis.[41,42] This end-to-end learning design is the fundamental basis of DL. In this field, the use of graphical processing units (GPUs) is widespread given their ability to parallelise the processing of large amounts of data.[43]

Image processing software

One of the most challenging problems that ML faces in radiology is the creation of large datasets of correctly labelled medical images.[7] Moreover, most of the image-based applications of AI require raw image processing before the training of the artificial neural network (ANN).[44,45] These algorithms require the resizing of input images so that they all have the same size, the reduction of noise in the images, and the normalisation of the images in order to map intensities of all images on a reference scale. In the case of MRI images, it is necessary to correct the bias field in order to correct image contrast variations due to magnetic field inhomogeneity.[46] As ML always operates in the presence of a supervisor or a teacher, supervised learning paradigms such as classification or segmentation tools, require close collaboration between radiologists and engineers to perform the correct segmentation of medical images in order to provide labelled data to ML models.[47]

One of the final goals of DL is to eliminate the dependence of post-processing data work for both radiologists and engineers; however, at this point, the help of engineers is essential in order to prepare medical images to feed into the ML algorithms and to select the most appropriate segmentation algorithms and applications for each imaging technique and anatomical area.[48] The most suitable segmentation algorithm for each case depends on the imaging technique or sequence, and on the anatomical area or lesion to be segmented (Fig 2).[49] It is desirable that the segmentation tool be as automatic as possible as manual segmentation is a very time-consuming task.[50] Nonetheless, the results of automatic segmentation should be reviewed by a radiologist. The time required for scrubbing, preprocessing and tagging images is often too large to develop ML algorithms in a reasonable timeframe. Therefore, when trying to bring AI to radiology a key theme is reducing this time by investing in highly scalable and parallelisable platforms that allow this process to be faster. This ensures the availability of a high amount of data prepared for the data scientist to be able to develop AI models within realistic timeframes. The investment in an efficient infrastructure or data lake able to retrieve data from PACS, integrate it and prepare these data for scrubbing, preprocessing and tagging in a highly parallelisable environment is an important expense.[51] Nevertheless, to be competitive, finance is crucial in the “big data” environment.

Statistical analysis

The application of complex statistical models for managing all the information derived from big data analysis is a field shared between radiologists, data scientists, and engineers. In some situations, radiologists require the collaboration of statisticians and engineers for interpretation of statistical results, especially to weigh up variables and recognise prediction errors due to bias or variance. Bias measures the difference between the prediction of the model and the correct value that it is trying to predict. Variance is the variability of model prediction for a given data point and it is an indicator that the model does not properly generalise new data. Minimising both bias and variance must be considered by radiologists when

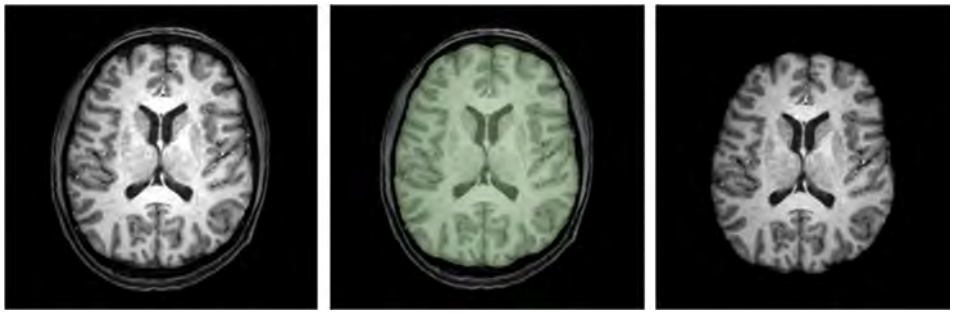


Figure 2 Skull stripped using a U-Net CNN implemented in DeepBrain (https://pypi.org/project/deepbrain/).

designing an ideal model.[52] Differentiation between statistics and ML algorithms is another concept that radiologist usually mix up, and whose objectives are quite different.[53] Regarding the purpose, statistical methods often focus on inference, which is reached through the creation and fitting of a model-based approach. Otherwise, ML usually concentrates on out-of-sample generalisation approaches, also known as extrapolation in classical statistics, and on finding patterns in copious hard-to-handle data.[54]

Natural language processing

Due to the unstructured nature of free-text radiology reports, their conversion into a computer-manageable representation is a challenge for radiologists. Natural language processing (NLP) is a field that aims to program machines to interpret language as humans do.[55] Data analysts and data scientists can help radiologists convert human-readable documents into structured computer data through NLP. In addition, NLP techniques enable automatic identification and extraction of information from the radiology reports.[56] Among other applications, NLP techniques can be used to identify pathologies in clinical reports, to index results in a searchable database, to provide patient- or report-level classification, to summarise findings in simpler natural language, or to assign protocols automatically to patient radiological examinations.[57,58] Moreover, the use of AI and NLP algorithms for text interpretation may help to boost the use of big data through the integration of radiological reports and images. With the progressive introduction of structured radiological reports, it is expected that the role of NPL will increase in the automation of multiple daily administrative tasks.

What engineers and data scientists need from radiologists

Set clinical targets

There are potential applications of AI algorithms in different aspects of radiology, ranging from study and workflow planification to imaging acquisition and post-processing, including applications facilitating lesion detection and segmentation or imaging registration. Most of these applications arise from radiologists' requirement to reduce time-consuming repetitive tasks that can be performed automatically, and even the potential desires to improve the radiologists' results in specific areas in which they have limited experience.[17] Furthermore, ML-based tools are now introducing new information in radiological reports, as algorithms can detect and quantify features that radiologists normally do not report or describe, in an unspecific manner[59] (Fig 3). Nowadays, the integration of clinical and image-based information into ML-based analysis is basic for engineers and data scientists. Most engineers do not understand the issues faced by radiologists in their daily routine practice and required radiologist input. By receiving this information, engineers should be able to start to develop potential applications based on ML or DL algorithms that make workflow easier, more accurate, and more efficient for radiologists. In the actual scenario, massive data alone will not provide any information unless there is a selection based on a strong and clinically proven hypothesis. Engineers and data scientists need the clinical view of radiologists to posit these hypotheses as radiologists prefer to reject a null hypothesis with a high level of significance rather than accept it, to explain that there is

Figure 3 Automatic detection and measurement of middle ascending and proximal arch thoracic dilatation using AI tools.

nothing clinically relevant present on the images.[60] For example, the need for automatic detection of brain haemorrhage is a clinical target to be translated to engineers and data scientists.

Imaging techniques and requirements

The wide range of imaging techniques that radiologists handle may be a drawback rather than an advantage for data scientists and engineers. Depending on the final target, some imaging techniques are more suitable than others for project development, and engineers need to know the advantages and disadvantages of each of these imaging techniques in order to choose the most appropriate one to obtain the most accurate results.[61] Moreover, there are specific ML and DL algorithms that will only work if they are applied using a certain imaging technique (i.e., detection of breast microcalcifications on mammography instead of breast MRI or automatic recognition of the pattern of breast tumour enhancement using MRI instead of computed tomography [CT]).[62] However, sometimes the best solution for extracting representative data from a certain structure or disease is a mix of modalities. Moreover, data derived in terms of signal intensity, Hounsfield units, or DICOM data, which will be used as raw data for statistical analysis or computed data, may vary depending on the imaging technique chosen.[36] Minimum requisites for image acquisition may also be provided to engineers to adjust software input data, obtain reproducible results, and detect potential sources of malfunction in the AI algorithms when real-world data are analysed (thickness, window level, gap, use of exogenous contrast agents, etc.).[1] For these reasons, radiologists should provide engineers with basic teaching points to select the most appropriate image technique according to the clinical target and the requirements, accuracy and computed features of the algorithm for the development of a specific AI application. In the example of detection of brain haemorrhage referred to above, the use of CT images should be the better option not only because the high contrast between acute bleeding and normal brain tissue, but also as this imaging technique is almost available worldwide.

Labelling data and ground truth

In order to train a model, data labelling/tagging is a crucial task for radiologists and the most important issue for engineers.[61] The existence of an accurate and balanced ground truth is the keystone for a model to really achieve the desired results.[63] Obtaining a quality ground truth is one of the major weaknesses for AI. If the ground truth is not appropriate, results from AI algorithms are going to be confusing and it will be impossible to reach valid conclusions. Among the main subtypes of ML, supervised algorithms can provide robust results with a lesser amount of data compared with non-supervised ML approaches.[64] For that purpose, it is necessary to label images for ML models that will be trained and learn to be able to discriminate between healthy and pathological tissues or between tissues with different properties. In this line, several groups have developed repositories of tagged images in different anatomical areas, such as brain atlases, basic for segmentation tasks.[65] Training a model with incorrect or poorly labelled data will lead to incorrect results. For these reasons, engineers need experienced radiologists for labelling both normal and pathological structures in medical images, which will provide high-quality data for further post-processing to obtain reliable and robust algorithms. For the example of brain haemorrhage detection, prior imaging labelling by radiologists of real cases of brain hemorrhage on CT for ML algorithms training is mandatory.

ML features selection to obtain preliminary models

During the pre-processing step, it is very important to select the most important variables in order to reduce the dimensionality of the data. Although advanced techniques that reduce such dimensionalities, such as principal component analysis or decision trees, can be applied, a prior variable selection must be carried out.[66] In this field, experienced radiologists may help engineers with the proposal and selection of the most representative clinical and radiological variables that may provide enough evidence to the results provided by ML or DL algorithms.[18] Data, such as patient age, specific metabolic, molecular or genetic biomarkers, histological subtype of neoplasia, anatomical landmarks, or number/distribution of lesions may not be relevant for engineers; however, these items make the difference for radiologists in clinical practice. For these reasons, improvement of communication and collaboration between radiologists and engineers may enhance the clinical applicability and the results of AI algorithms, as they will be focused on solving real radiological problems from the design phase. In the case of brain haemorrhage detection, radiologists may guide engineers for identifying potential features (i.e., radiodensity) that ML algorithms could use to determine the presence of acute, subacute, or chronic bleeding.

Basic anatomy and physiopathology

In a near future ML and DL systems will became a reality in radiology, supporting, and boosting the diagnostic capacity of radiologists and extracting prognostic information from imaging.[67–69] In order to create tools that follow radiological protocols, ML and DL applications require knowledge of anatomy and physiopathology on which the training algorithm is focused. Moreover, to establish the possible approaches and interpretation of the results, basic concepts regarding human anatomy and physiopathology are needed to obtain an initial approximation of the biological likelihood of the results obtained.[70] It is well known by medical and radiological communities that specific organs or human systems host specific diseases within specific characteristics. Thus, results provided by AI algorithms should fit into the expected results for each AI task.[4] In this field, radiologists should help guide engineers to boost AI algorithms with prior medical knowledge and avoid post-

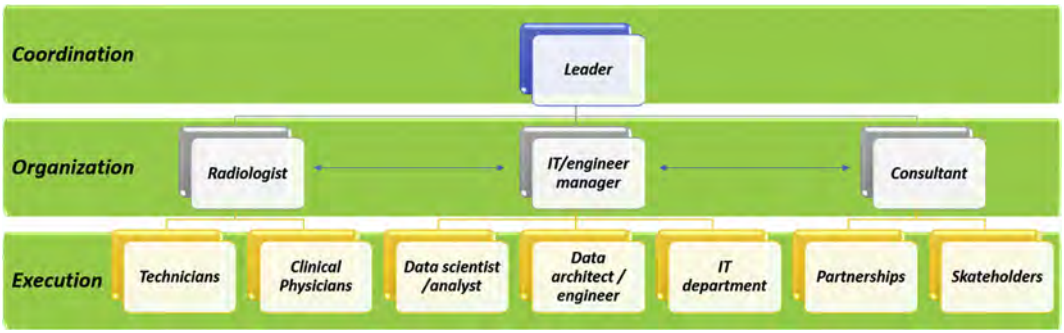


Figure 4 Proposal for hierarchical organisation of an AI multidisciplinary team to develop integral radiological solutions.

processes that provide unrealistic data. This step will help avoid the misinterpretation of classical statistical inference as well as remove noise sources or acquisition failures that may hinder the overall accuracy of AI algorithms generated. For the example of brain haemorrhage detection, the importance of selection of anatomical region (head) as well as location or volume of brain haemorrhage, are characteristics that should be provided in detail to engineers to optimise the clinical relevance of ML algorithms.

How to build a successful multidisciplinary AI team

A multidisciplinary approach will power the development and application of AI solutions into real clinical situations.[71] A standard AI team should comprise radiologists, engineers (including bioinformatics), data architects, data analysts, and data scientists. It is mandatory to have a qualified information technology (IT) department for technical and logistical support related to storage (physical or in-cloud), computer servicing, handling security issues, and data exchange. The inclusion of technicians or radiographers to perform non-supervised or semi-supervised routine tasks may save time and aid with training and teaching subspecialised technicians in the AI environment. Sporadic or continuous contact and feedback from clinical physicians may be useful for identifying specific targets outside the radiological perspective enhancing the final value of the potential AI tool developed; however, other players, such as consultants to update policies and regulatory issues related to patient data protection, potential stakeholders (with special interests in academy environment and vendors), and even clinical trials for testing AI prototypes are essential for performing the non-scientific aspects of the work.[72,73] Above all, a leadership figure must be clearly identified to guide the team and identify the priorities of the group. In this line, a clear target must be defined focusing on a relevant issue for clinical and radiological practice. This target may range from image analysis focused on radiological interpretation and reporting (the most common branch of AI exploited by far) to specific reconstructions or pre- and post-processing tools.[74,75] Studies protocolisation, handling of patient schedules, worklist prioritisation as well as other administrative and management task will be suitable for AI algorithms developments (Fig 4).[76–78] Another key factor for multidisciplinary team performance is to maintain close and continuous communication between all members of the group. In these periodical meetings, continuous feedback and reporting of potential issues or drawbacks in the development and implementation of AI tools must be clarified. Doubts and unexpected requirements from both radiologists and engineers should be exposed and taken into consideration to improve the final product and correct it to solve or help with real radiological or administrative tasks. For example, it is very important for the group to agree upon and decide whether or not the AI prototype is going to be integrated automatically for image analysis of all patients, or if it is going to be implemented on-demand at the request of radiologists only in certain scenarios.

Conclusions

AI-based tools are being used in daily clinical radiological practice and undoubtedly, in the more immediate future, their role will increase. The validation of existing AI applications and the development of new AI-based solutions for common radiological problems and researching projects depends on the collaborative work between radiologists, data scientists, and engineers. The scarce specialised training of radiologists in AI algorithms and the understandable lack of knowledge of engineers regarding biomedical and radiological concepts, justify the need to build multidisciplinary teams that include all specialists. Knowledge of what they need from each other will reduce this gap, improve communication, and help to reach more precise, useful, and innovative AI applications in the field of radiology.

Declarations of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Antonio Luna is occasional lecturer of Philips, Siemens Healthineers, Bracco and Canon and receives royalties as book editor from Springer-Verlag. Dr. Martín-Noguerol from MRI unit, Radiology department. HT medica, Jaén, (SPAIN) has nothing to disclose. Félix Paulano-

Godino from Engineering department. HT medica. Jaén, (SPAIN) has nothing to disclose. Rafael López-Ortega from Engineering department. HT medica. Jaén (SPAIN) has nothing to disclose.

Acknowledgements

The authors thank Camilo Riascos for revision of the final manuscript draft.

References

1. Hosny A, Parmar C, Quackenbush J, *et al*. Artificial intelligence in radiology. *Nat Rev Canc* 2018:1–11. https://doi.org/10.1038/s41568-018-0016-5.
2. Saba L, Biswas M, Kuppili V, *et al*. The present and future of deep learning in radiology. *Eur J Radiol* 2019;**114**:14–24. https://doi.org/10.1016/j.ejrad.2019.02.038.
3. Recht M, Nick Bryan R, Recht MP. Artificial Intelligence: threat or boon to radiologists? *J Am Coll Radiol* 2017;**14**:1476–80. https://doi.org/10.1016/j.jacr.2017.07.007.
4. Oakden-Rayner L. The rebirth of CAD: how is modern ai different from the CAD we know? *Radiol Artif Intell* 2019;**1**(3):e180089, https://doi.org/10.1148/ryai.2019180089.
5. Shah P, Kendall F, Khozin S, *et al*. Artificial intelligence and machine learning in clinical development: a translational perspective. *Npj Digit Med* 2019;**2**:69. https://doi.org/10.1038/s41746-019-0148-3.
6. Ghesu FC, Georgescu B, Grbic S, *et al*. Towards intelligent robust detection of anatomical structures in incomplete volumetric data. *Med Image Anal* 2018;**48**:203–13. https://doi.org/10.1016/j.media.2018.06.007.
7. European Society of Radiology. What the radiologist should know about artificial intelligence — an ESR white paper. *Insights Imaging* 2019;**10**:44. https://doi.org/10.1186/s13244-019-0738-2.
8. Sabottke CF, Spieler BM. The effect of image resolution on deep learning in radiography. *Radiol Artif Intell* 2020;**2**:e190015, https://doi.org/10.1148/ryai.2019190015.
9. Adadi A, Berrada M. Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE Access* 2018;**6**:52138–60. https://doi.org/10.1109/ACCESS.2018.2870052.
10. Rubin DL. Artificial intelligence in imaging: the radiologist's role. *J Am Coll Radiol* 2019;**16**:1309–17. https://doi.org/10.1016/j.jacr.2019.05.036.
11. Holzinger A, Langs G, Denk H, *et al*. *Causability and explainability of artificial intelligence in medicine***9**. Wiley Interdiscip Rev Data Min Knowl Discov; 2019e1312. https://doi.org/10.1002/widm.1312. Epub 2019 Apr 2.
12. Erickson BJ. Magician's corner: how to start learning about deep learning. *Radiol Artif Intell* 2019;**1 x190072**, https://doi.org/10.1148/ryai.2019190072.
13. Sogani J, Allen B, Dreyer K, *et al*. Artificial intelligence in radiology: the ecosystem essential to improving patient care. *Clin Imag* 2020 Jan;**59**(1):A3–6. https://doi.org/10.1016/j.clinimag.2019.08.001.
14. Allen B, Gish R, Dreyer K. *The role of an artificial intelligence ecosystem in radiology*. *Artificial intelligence in medical imaging: opportunities, applications and risks*. Cham: Springer International Publishing; 2019. p. 291–327. https://doi.org/10.1007/978-3-319-94878-2_19.
15. Liu S, Wang X, Liu M, *et al*. Towards better analysis of machine learning models: a visual analytics perspective. *Vis Inform* 2017;**1**:48–56. https://doi.org/10.1016/j.visinf.2017.01.006.
16. Hohman F, Kahng M, Pienta R, *et al*. Visual analytics in deep learning: an interrogative survey for the next frontiers. *IEEE Trans Vis Comput Graph* 2018;**4**, https://doi.org/10.1109/TVCG.2018.2843369. 10.1109/TVCG.2018.2843369.
17. Thrall JH, Li X, Li Q, *et al*. Artificial intelligence and machine learning in radiology: opportunities, challenges, pitfalls, and criteria for success. *J Am Coll Radiol* 2018;**15**:504–8. https://doi.org/10.1016/j.jacr.2017.12.026.
18. Chartrand G, Cheng PM, Vorontsov E, *et al*. Deep learning: a primer for radiologists. *RadioGraphics* 2017;**37**:2113–31. https://doi.org/10.1148/rg.2017170077.
19. Baselli G, Codari M, Sardanelli F. Opening the black box of machine learning in radiology: can the proximity of annotated cases be a way? *Eur Radiol Exp* 2020;**4**:30. https://doi.org/10.1186/s41747-020-00159-0.
20. Handelman GS, Kok HK, Chandra RV, *et al*. Peering into the black box of artificial intelligence: evaluation metrics of machine learning methods. *AJR Am J Roentgenol* 2019 Jan;**212**(1):38–43. https://doi.org/10.2214/AJR.18.20224.
21. Zeiler MD, Fergus R. Visualizing and understanding convolutional networks. In: *Lecture Notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)*. Cham: Springer Verlag; 2014. p. 818–33.
22. Simonyan K, Vedaldi A, Zisserman A. Deep inside convolutional networks: visualising image classification models and saliency maps. In: *2nd international conference on learning representations, ICLR 2014 — workshop track proceedings*. Canada: Banff, AB; 2014. p. 14–6. 19 April, https://dblp.org/rec/journals/corr/SimonyanVZ13.
23. Khedher L, Illán IA, Górriz JM, *et al*. Independent component analysis-support vector machine-based computer-aided diagnosis system for Alzheimer's with visual support. *Int J Neural Syst* 2017;**27**:1650050. https://doi.org/10.1142/S0129065716500507.
24. Wongsuphasawat K, Smilkov D, Wexler J, *et al*. Visualizing dataflow graphs of deep learning models in TensorFlow. *IEEE Trans Vis Comput Graph* 2018;**24**:1–12. https://doi.org/10.1109/TVCG.2017.2744878.
25. Kahng M, Andrews PY, Kalro A, Chau DHP. ActiVis: Visual Exploration of Industry-Scale Deep Neural Network Models. *IEEE Trans Vis Comput Graph* 2018;**24**(1):88–97. https://doi.org/10.1109/TVCG.2017.2744718.
26. Barredo Arrieta A, Díaz-Rodríguez N, Del Ser J, *et al*. Explainable Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf Fusion* 2020;**58**:82–115. https://doi.org/10.1016/j.inffus.2019.12.012.
27. Zhang D, Shen D. Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease. *Neuroimage* 2012;**59**:895–907. https://doi.org/10.1016/j.neuroimage.2011.09.069.
28. Gong H, Yu L, Leng S, *et al*. A deep learning- and partial least square regression-based model observer for a low-contrast lesion detection task in CT. *Med Phys* 2019;**46**:2052–63. https://doi.org/10.1002/mp.13500.
29. Zhang S, Zhang C, Yang Q. Data preparation for data mining. *Appl Artif Intell* 2003 May;**17**:375–81. https://doi.org/10.1080/08839510390219264.
30. Yamashita R, Nishio M, Kinh R, *et al*. Convolutional neural networks: an overview and application in radiology. *Insights Imaging* 2018;**9**:611–29. https://doi.org/10.1007/s13244-018-0639-9.
31. Morrison JJ, Hostetter J, Wang K, *et al*. Data-driven decision support for radiologists: re-using the national lung screening trial dataset for pulmonary nodule management. *J Digit Imag* 2014;**28**:18–23. https://doi.org/10.1007/s10278-014-9720-1.
32. Moore SM, Maffitt DR, Smith KE, *et al*. De-identification of medical images with retention of scientific research value. *RadioGraphics* 2015;**35**:727–35. https://doi.org/10.1148/rg.2015140244.
33. van Ooijen PMA. Quality and curation of medical images and data. In: *Artificial intelligence in medical imaging*. Cham: Springer International Publishing; 2019. p. 247–55. https://doi.org/10.1007/978-3-319-94878-2_17.
34. Wu B, Wang C, Yao H. Security analysis and secure channel-free certificate less searchable public key authenticated encryption for a cloud-based Internet of things. *PLoS One* 2020;**15**:e0230722, https://doi.org/10.1371/journal.pone.0230722.
35. Rockall A. From hype to hope to hard work: developing responsible AI for radiology. *Clin Radiol* 2020;**75**:1–2. https://doi.org/10.1016/j.crad.2019.09.123.
36. Riddle WR, Pickens DR. Extracting data from a DICOM file. *Med Phys* 2005;**32**:1537–41. https://doi.org/10.1118/1.1916183.
37. Aryanto KYE, Oudkerk M, van Ooijen PMA. Free DICOM de-identification tools in clinical research: functioning and safety of patient privacy. *Eur Radiol* 2015 Dec;**25**:3685–95. https://doi.org/10.1007/s00330-015-3794-0.
38. Juntu J, Sijbers J, Dyck D, *et al*. Bias field correction for MRI images. In: Kurzyński M, Puchała E, Woźniak M, *et al*., editors. *Computer recognition systems. Advances in soft computing*vol. 30. Berlin: Springer; 2005. 543–51. https://doi.org/10.1007/3-540-32390-2_64.
39. Yasaka K, Akai H, Kunimatsu A, *et al*. Deep learning with convolutional neural network in radiology. *Jpn J Radiol* 2018;**36**:257–72. https://doi.org/10.1007/s11604-018-0726-3.

40. Sessions V, Valtorta M. The effects of data quality on machine learning algorithms. In: *Proceedings of the 2006 international conference on information quality*vol. 2006. ICIQ; 2006. p. 485–98.
41. Wainberg M, Merico D, Delong A, *et al*. Deep learning in biomedicine. *Nat Biotechnol* 2018;**36**:829–38. https://doi.org/10.1038/nbt.4233.
42. Ather S, Kadir T, Gleeson F. Artificial intelligence and radiomics in pulmonary nodule management: current status and future applications. *Clin Radiol* 2020;**75**:13–9. https://doi.org/10.1016/j.crad.2019.04.017.
43. Steinkraus D, Buck I, Simard PY. Using GPUs for machine learning algorithms. In: *Proceedings of the international conference on document analysis and recognition*. ICDAR; 2005, https://doi.org/10.1109/ICDAR.2005.251.
44. Seo S, Do WJ, Luu HM, *et al*. Artificial neural network for slice encoding for metal artifact correction (SEMAC) MRI. *Magn Reson Med* 2020;**84**:263–76. https://doi.org/10.1002/mrm.28126.
45. Alis D, Bagcilar O, Senli YD, *et al*. The diagnostic value of quantitative texture analysis of conventional MRI sequences using artificial neural networks in grading gliomas. *Clin Radiol* 2020;**75**:351–7. https://doi.org/10.1016/j.crad.2019.12.008.
46. Akkus Z, Galimzianova A, Hoogi A, *et al*. Deep learning for brain MRI segmentation: state of the art and future directions. *J Digit Imag* 2017;**30**:449–59. https://doi.org/10.1007/s10278-017-9983-4.
47. Jena M, Prava Mishra S, Mishra D. A survey on applications of machine learning techniques for medical image segmentation. *Artic Int J Eng Technol* 2018;**7**(4):4489–95. https://doi.org/10.14419/ijet.v7i4.19005.
48. Lecun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;**521**(7553):436–44. https://doi.org/10.1038/nature14539.
49. Ronneberger O, Fischer P, Brox T. *U-net: Convolutional networks for biomedical image segmentation*. In: *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol 9351; 2015. p. 234–41. https://doi.org/10.1007/978-3-319-24574-4_28.
50. Hu P, Wu F, Peng J, *et al*. Automatic abdominal multi-organ segmentation using deep convolutional neural network and time-implicit level sets. *Int J Comput Assist Radiol Surg* 2017;**12**:399–411. https://doi.org/10.1007/s11548-016-1501-5.
51. O'Leary DE. Embedding AI and crowdsourcing in the big data lake. *IEEE Intell Syst* 2014 Sep 1;**29**(5):70–3. https://doi.org/10.1109/MIS.2014.82.
52. Kong EB, Dietterich TG. *Error-correcting output coding corrects bias and variance*. *Machine learning proceedings 1995*. London: Elsevier; 1995. p. 313–21. https://doi.org/10.1016/B978-1-55860-377-6.50046-3.
53. Bzdok D, Altman N, Krzywinski M. *Points of significance: statistics versus machine learning*. London: Nature Publishing Group; 2018. p. 1–7.
54. Yeung DY, Chang H, Dai G. A scalable kernel-based semisupervised metric learning algorithm with out-of-sample generalization ability. *Neural Comput* 2008;**20**:2839–61. https://doi.org/10.1162/neco.2008.05-07-528.
55. Huesch MD, Cherian R, Labib S, *et al*. Evaluating report text variation and informativeness: natural language processing of CT chest imaging for pulmonary embolism. *J Am Coll Radiol* 2018;**15**:554–62. https://doi.org/10.1016/j.jacr.2017.12.017.
56. Yetisgen-Yildiz M, Gunn ML, Xia F, *et al*. A text processing pipeline to extract recommendations from radiology reports. *J Biomed Inform* 2013;**46**:354–62. https://doi.org/10.1016/j.jbi.2012.12.005.
57. Jungmann F, Arnhold G, Kämpgen B, *et al*. A hybrid reporting platform for extended RadLex coding combining structured reporting templates and natural language processing. *J Digit Imag* 2020, https://doi.org/10.1007/s10278-020-00342-0.
58. Pons E, Braun LMM, Hunink MGM, *et al*. Natural language processing in radiology: a systematic review. *Radiology* 2016;**279**:329–43. https://doi.org/10.1148/radiol.16142770.
59. Langs G, R€ohrich S, Hofmanninger J, *et al*. Machine learning: from radiomics to discovery and routine. *Radiologe* 2018;**58**:1–6. https://doi.org/10.1007/s00117-018-0407-3.
60. Friston K. *Ten Ironic Rules for Non-statistical Reviewers Neuroimage* 2012;**61**:1300–10. https://doi.org/10.1016/j.neuroimage.2012.04.018.
61. Choy G, Khalilzadeh O, Michalski M, *et al*. Current applications and future impact of machine learning in radiology. *Radiology* 2018;**288**:318–28 http://pubs.rsna.org/doi/10.1148/radiol.2018171820.
62. Samala RK, Chan HP, Lu Y, *et al*. Digital breast tomosynthesis: computer-aided detection of clustered microcalcifications on planar projection images. *Phys Med Biol* 2014;**59**:7457–77. https://doi.org/10.1088/0031-9155/59/23/7457.
63. Martín Noguerol T, Paulano-Godino F, Martín-Valdivia MT, *et al*. Strengths, weaknesses, opportunities, and threats analysis of artificial intelligence and machine learning applications in radiology. *J Am Coll Radiol* 2019;**16**:1239–47. https://doi.org/10.1016/j.jacr.2019.05.047.
64. Baştanlar Y, Özuysal M. Introduction to machine learning. *Methods Mol Biol* 2014;**1107**:105–28 http://link.springer.com/10.1007/978-1-62703-748-8_7.
65. Iqbal A, Khan R, Karayannis T. Developing a brain atlas through deep learning. *Nat Mach Intell* 2019 Jun 10;**1**(6):277–87. https://doi.org/10.1038/s42256-019-0058-8.
66. Cao LJ, Chua KS, Chong WK, *et al*. A comparison of PCA, KPCA and ICA for dimensionality reduction in support vector machine. *Neurocomputing* 2003;**55**:321–36. https://doi.org/10.1016/S0925-2312(03)00433-8.
67. Liew C. The future of radiology augmented with artificial intelligence: a strategy for success. *Eur J Radiol* 2018;**102**:152–6. https://doi.org/10.1016/j.ejrad.2018.03.019.
68. Jha S, Topol EJ. Adapting to artificial intelligence: radiologists and pathologists as information specialists. *JAMA* 2016;**316**:2353–4. https://doi.org/10.1001/jama.2016.17438.
69. Korfiatis P, Erickson B. Deep learning can see the unseeable: predicting molecular markers from MRI of brain gliomas. *Clin Radiol* 2019;**74**(5):367–73. https://doi.org/10.1016/j.crad.2019.01.028.
70. Kannampallil TG, Franklin A, Mishra R, *et al*. Understanding the nature of information seeking behavior in critical care: implications for the design of health information technology. *Artif Intell Med* 2013;**57**:21–9. https://doi.org/10.1016/j.artmed.2012.10.002.
71. Di Ieva A. AI-augmented multidisciplinary teams: hype or hope? *Lancet* 2019;**394**(10211):1801. https://doi.org/10.1016/S0140-6736(19)32626-1.
72. Dwivedi YK, Hughes L, Ismagilova E, *et al*. Artificial intelligence (AI): multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *Int J Inf Manage* 2019;**101994** https://linkinghub.elsevier.com/retrieve/pii/S026840121930917X.
73. Gilbert FJ, Smye SW, Schönlieb CB. Artificial intelligence in clinical imaging: a health system approach. *Clin Radiol* 2020;**75**:3–6. https://doi.org/10.1016/j.crad.2019.09.122.
74. Yasaka K, Abe O. Deep learning and artificial intelligence in radiology: current applications and future directions. *PLOS Med* 2018;**15**:e1002707 https://dx.plos.org/10.1371/journal.pmed.1002707.
75. Codari M, Melazzini L, Morozov SP, *et al*. Impact of artificial intelligence on radiology: a EuroAIM survey among members of the European Society of Radiology. *Insights Imaging* 2019;**10**:105 https://insightsimaging.springeropen.com/articles/10.1186/s13244-019-0798-3.
76. Winkel DJ, Heye T, Weikert TJ, *et al*. Evaluation of an AI-based detection software for acute findings in abdominal computed tomography scans: toward an automated work list prioritization of routine CT examinations. *Invest Radiol* 2019;**54**:55–9. https://doi.org/10.1097/RLI.0000000000000509.
77. Prevedello LM, Erdal BS, Ryu JL, *et al*. Automated critical test findings identification and online notification system using artificial intelligence in imaging. *Radiology* 2017;**285**:923–31. https://doi.org/10.1148/radiol.2017162664.
78. Kuo W, Häne C, Mukherjee P, *et al*. Expert-level detection of acute intracranial hemorrhage on head computed tomography using deep learning. *Proc Natl Acad Sci U S A* 2019;**116**:22737–45. https://doi.org/10.1073/pnas.1908021116.

EXHIBIT R-3

Rich feature hierarchies for accurate object detection and semantic segmentation

Tech report (v5)

Ross Girshick Jeff Donahue Trevor Darrell Jitendra Malik
UC Berkeley
{rbg,jdonahue,trevor,malik}@eecs.berkeley.edu

Abstract

Object detection performance, as measured on the canonical PASCAL VOC dataset, has plateaued in the last few years. The best-performing methods are complex ensemble systems that typically combine multiple low-level image features with high-level context. In this paper, we propose a simple and scalable detection algorithm that improves mean average precision (mAP) by more than 30% relative to the previous best result on VOC 2012—achieving a mAP of 53.3%. Our approach combines two key insights: (1) one can apply high-capacity convolutional neural networks (CNNs) to bottom-up region proposals in order to localize and segment objects and (2) when labeled training data is scarce, supervised pre-training for an auxiliary task, followed by domain-specific fine-tuning, yields a significant performance boost. Since we combine region proposals with CNNs, we call our method R-CNN: *Regions with CNN features. We also compare R-CNN to OverFeat, a recently proposed sliding-window detector based on a similar CNN architecture. We find that R-CNN outperforms OverFeat by a large margin on the 200-class ILSVRC2013 detection dataset. Source code for the complete system is available at* http://www.cs.berkeley.edu/~rbg/rcnn.

1. Introduction

Features matter. The last decade of progress on various visual recognition tasks has been based considerably on the use of SIFT [29] and HOG [7]. But if we look at performance on the canonical visual recognition task, PASCAL VOC object detection [15], it is generally acknowledged that progress has been slow during 2010-2012, with small gains obtained by building ensemble systems and employing minor variants of successful methods.

SIFT and HOG are blockwise orientation histograms, a representation we could associate roughly with complex cells in V1, the first cortical area in the primate visual pathway. But we also know that recognition occurs several stages downstream, which suggests that there might be hierarchical, multi-stage processes for computing features that are even more informative for visual recognition.

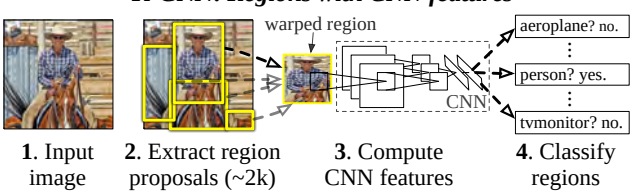


Figure 1: Object detection system overview. Our system (1) takes an input image, (2) extracts around 2000 bottom-up region proposals, (3) computes features for each proposal using a large convolutional neural network (CNN), and then (4) classifies each region using class-specific linear SVMs. R-CNN achieves a mean average precision (mAP) of **53.7% on PASCAL VOC 2010**. For comparison, [39] reports 35.1% mAP using the same region proposals, but with a spatial pyramid and bag-of-visual-words approach. The popular deformable part models perform at 33.4%. On the 200-class **ILSVRC2013 detection dataset, R-CNN's mAP is 31.4%**, a large improvement over OverFeat [34], which had the previous best result at 24.3%.

Fukushima's "neocognitron" [19], a biologically-inspired hierarchical and shift-invariant model for pattern recognition, was an early attempt at just such a process. The neocognitron, however, lacked a supervised training algorithm. Building on Rumelhart et al. [33], LeCun et al. [26] showed that stochastic gradient descent via back-propagation was effective for training convolutional neural networks (CNNs), a class of models that extend the neocognitron.

CNNs saw heavy use in the 1990s (e.g., [27]), but then fell out of fashion with the rise of support vector machines. In 2012, Krizhevsky et al. [25] rekindled interest in CNNs by showing substantially higher image classification accuracy on the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [9, 10]. Their success resulted from training a large CNN on 1.2 million labeled images, together with a few twists on LeCun's CNN (e.g., $\max(x, 0)$ rectifying non-linearities and "dropout" regularization).

The significance of the ImageNet result was vigorously

debated during the ILSVRC 2012 workshop. The central issue can be distilled to the following: To what extent do the CNN classification results on ImageNet generalize to object detection results on the PASCAL VOC Challenge?

We answer this question by bridging the gap between image classification and object detection. This paper is the first to show that a CNN can lead to dramatically higher object detection performance on PASCAL VOC as compared to systems based on simpler HOG-like features. To achieve this result, we focused on two problems: localizing objects with a deep network and training a high-capacity model with only a small quantity of annotated detection data.

Unlike image classification, detection requires localizing (likely many) objects within an image. One approach frames localization as a regression problem. However, work from Szegedy et al. [38], concurrent with our own, indicates that this strategy may not fare well in practice (they report a mAP of 30.5% on VOC 2007 compared to the 58.5% achieved by our method). An alternative is to build a sliding-window detector. CNNs have been used in this way for at least two decades, typically on constrained object categories, such as faces [32, 40] and pedestrians [35]. In order to maintain high spatial resolution, these CNNs typically only have two convolutional and pooling layers. We also considered adopting a sliding-window approach. However, units high up in our network, which has five convolutional layers, have very large receptive fields (195×195 pixels) and strides (32×32 pixels) in the input image, which makes precise localization within the sliding-window paradigm an open technical challenge.

Instead, we solve the CNN localization problem by operating within the "recognition using regions" paradigm [21], which has been successful for both object detection [39] and semantic segmentation [5]. At test time, our method generates around 2000 category-independent region proposals for the input image, extracts a fixed-length feature vector from each proposal using a CNN, and then classifies each region with category-specific linear SVMs. We use a simple technique (affine image warping) to compute a fixed-size CNN input from each region proposal, regardless of the region's shape. Figure 1 presents an overview of our method and highlights some of our results. Since our system combines region proposals with CNNs, we dub the method R-CNN: Regions with CNN features.

In this updated version of this paper, we provide a head-to-head comparison of R-CNN and the recently proposed OverFeat [34] detection system by running R-CNN on the 200-class ILSVRC2013 detection dataset. OverFeat uses a sliding-window CNN for detection and until now was the best performing method on ILSVRC2013 detection. We show that R-CNN significantly outperforms OverFeat, with a mAP of 31.4% versus 24.3%.

A second challenge faced in detection is that labeled data is scarce and the amount currently available is insufficient for training a large CNN. The conventional solution to this problem is to use *unsupervised* pre-training, followed by supervised fine-tuning (e.g., [35]). The second principle contribution of this paper is to show that *supervised* pre-training on a large auxiliary dataset (ILSVRC), followed by domain-specific fine-tuning on a small dataset (PASCAL), is an effective paradigm for learning high-capacity CNNs when data is scarce. In our experiments, fine-tuning for detection improves mAP performance by 8 percentage points. After fine-tuning, our system achieves a mAP of 54% on VOC 2010 compared to 33% for the highly-tuned, HOG-based deformable part model (DPM) [17, 20]. We also point readers to contemporaneous work by Donahue et al. [12], who show that Krizhevsky's CNN can be used (without fine-tuning) as a blackbox feature extractor, yielding excellent performance on several recognition tasks including scene classification, fine-grained sub-categorization, and domain adaptation.

Our system is also quite efficient. The only class-specific computations are a reasonably small matrix-vector product and greedy non-maximum suppression. This computational property follows from features that are shared across all categories and that are also two orders of magnitude lower-dimensional than previously used region features (*cf.* [39]).

Understanding the failure modes of our approach is also critical for improving it, and so we report results from the detection analysis tool of Hoiem et al. [23]. As an immediate consequence of this analysis, we demonstrate that a simple bounding-box regression method significantly reduces mislocalizations, which are the dominant error mode.

Before developing technical details, we note that because R-CNN operates on regions it is natural to extend it to the task of semantic segmentation. With minor modifications, we also achieve competitive results on the PASCAL VOC segmentation task, with an average segmentation accuracy of 47.9% on the VOC 2011 test set.

2. Object detection with R-CNN

Our object detection system consists of three modules. The first generates category-independent region proposals. These proposals define the set of candidate detections available to our detector. The second module is a large convolutional neural network that extracts a fixed-length feature vector from each region. The third module is a set of class-specific linear SVMs. In this section, we present our design decisions for each module, describe their test-time usage, detail how their parameters are learned, and show detection results on PASCAL VOC 2010-12 and on ILSVRC2013.

2.1. Module design

Region proposals. A variety of recent papers offer methods for generating category-independent region proposals.

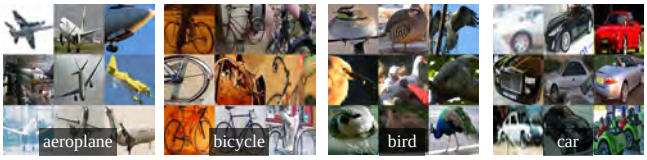


Figure 2: Warped training samples from VOC 2007 train.

Examples include: objectness [1], selective search [39], category-independent object proposals [14], constrained parametric min-cuts (CPMC) [5], multi-scale combinatorial grouping [3], and Cireşan et al. [6], who detect mitotic cells by applying a CNN to regularly-spaced square crops, which are a special case of region proposals. While R-CNN is agnostic to the particular region proposal method, we use selective search to enable a controlled comparison with prior detection work (e.g., [39, 41]).

Feature extraction. We extract a 4096-dimensional feature vector from each region proposal using the Caffe [24] implementation of the CNN described by Krizhevsky et al. [25]. Features are computed by forward propagating a mean-subtracted 227×227 RGB image through five convolutional layers and two fully connected layers. We refer readers to [24, 25] for more network architecture details.

In order to compute features for a region proposal, we must first convert the image data in that region into a form that is compatible with the CNN (its architecture requires inputs of a fixed 227×227 pixel size). Of the many possible transformations of our arbitrary-shaped regions, we opt for the simplest. Regardless of the size or aspect ratio of the candidate region, we warp all pixels in a tight bounding box around it to the required size. Prior to warping, we dilate the tight bounding box so that at the warped size there are exactly p pixels of warped image context around the original box (we use $p = 16$). Figure 2 shows a random sampling of warped training regions. Alternatives to warping are discussed in Appendix A.

2.2. Test-time detection

At test time, we run selective search on the test image to extract around 2000 region proposals (we use selective search's "fast mode" in all experiments). We warp each proposal and forward propagate it through the CNN in order to compute features. Then, for each class, we score each extracted feature vector using the SVM trained for that class. Given all scored regions in an image, we apply a greedy non-maximum suppression (for each class independently) that rejects a region if it has an intersection-over-union (IoU) overlap with a higher scoring selected region larger than a learned threshold.

Run-time analysis. Two properties make detection efficient. First, all CNN parameters are shared across all categories. Second, the feature vectors computed by the CNN are low-dimensional when compared to other common approaches, such as spatial pyramids with bag-of-visual-word encodings. The features used in the UVA detection system [39], for example, are two orders of magnitude larger than ours (360k vs. 4k-dimensional).

The result of such sharing is that the time spent computing region proposals and features (13s/image on a GPU or 53s/image on a CPU) is amortized over all classes. The only class-specific computations are dot products between features and SVM weights and non-maximum suppression. In practice, all dot products for an image are batched into a single matrix-matrix product. The feature matrix is typically 2000×4096 and the SVM weight matrix is $4096 \times N$, where N is the number of classes.

This analysis shows that R-CNN can scale to thousands of object classes without resorting to approximate techniques, such as hashing. Even if there were 100k classes, the resulting matrix multiplication takes only 10 seconds on a modern multi-core CPU. This efficiency is not merely the result of using region proposals and shared features. The UVA system, due to its high-dimensional features, would be two orders of magnitude slower while requiring 134GB of memory just to store 100k linear predictors, compared to just 1.5GB for our lower-dimensional features.

It is also interesting to contrast R-CNN with the recent work from Dean et al. on scalable detection using DPMs and hashing [8]. They report a mAP of around 16% on VOC 2007 at a run-time of 5 minutes per image when introducing 10k distractor classes. With our approach, 10k detectors can run in about a minute on a CPU, and because no approximations are made mAP would remain at 59% (Section 3.2).

2.3. Training

Supervised pre-training. We discriminatively pre-trained the CNN on a large auxiliary dataset (ILSVRC2012 classification) using *image-level annotations* only (bounding-box labels are not available for this data). Pre-training was performed using the open source Caffe CNN library [24]. In brief, our CNN nearly matches the performance of Krizhevsky et al. [25], obtaining a top-1 error rate 2.2 percentage points higher on the ILSVRC2012 classification validation set. This discrepancy is due to simplifications in the training process.

Domain-specific fine-tuning. To adapt our CNN to the new task (detection) and the new domain (warped proposal windows), we continue stochastic gradient descent (SGD) training of the CNN parameters using only warped region proposals. Aside from replacing the CNN's ImageNet-specific 1000-way classification layer with a randomly initialized $(N + 1)$-way classification layer (where N is the number of object classes, plus 1 for background), the CNN architecture is unchanged. For VOC, $N = 20$ and for ILSVRC2013, $N = 200$. We treat all region proposals with

≥ 0.5 IoU overlap with a ground-truth box as positives for that box's class and the rest as negatives. We start SGD at a learning rate of 0.001 (1/10th of the initial pre-training rate), which allows fine-tuning to make progress while not clobbering the initialization. In each SGD iteration, we uniformly sample 32 positive windows (over all classes) and 96 background windows to construct a mini-batch of size 128. We bias the sampling towards positive windows because they are extremely rare compared to background.

Object category classifiers. Consider training a binary classifier to detect cars. It's clear that an image region tightly enclosing a car should be a positive example. Similarly, it's clear that a background region, which has nothing to do with cars, should be a negative example. Less clear is how to label a region that partially overlaps a car. We resolve this issue with an IoU overlap threshold, below which regions are defined as negatives. The overlap threshold, 0.3, was selected by a grid search over $\{0, 0.1, \ldots, 0.5\}$ on a validation set. We found that selecting this threshold carefully is important. Setting it to 0.5, as in [39], decreased mAP by 5 points. Similarly, setting it to 0 decreased mAP by 4 points. Positive examples are defined simply to be the ground-truth bounding boxes for each class.

Once features are extracted and training labels are applied, we optimize one linear SVM per class. Since the training data is too large to fit in memory, we adopt the standard hard negative mining method [17, 37]. Hard negative mining converges quickly and in practice mAP stops increasing after only a single pass over all images.

In Appendix B we discuss why the positive and negative examples are defined differently in fine-tuning versus SVM training. We also discuss the trade-offs involved in training detection SVMs rather than simply using the outputs from the final softmax layer of the fine-tuned CNN.

2.4. Results on PASCAL VOC 2010-12

Following the PASCAL VOC best practices [15], we validated all design decisions and hyperparameters on the VOC 2007 dataset (Section 3.2). For final results on the VOC 2010-12 datasets, we fine-tuned the CNN on VOC 2012 train and optimized our detection SVMs on VOC 2012 trainval. We submitted test results to the evaluation server only once for each of the two major algorithm variants (with and without bounding-box regression).

Table 1 shows complete results on VOC 2010. We compare our method against four strong baselines, including SegDPM [18], which combines DPM detectors with the output of a semantic segmentation system [4] and uses additional inter-detector context and image-classifier rescoring. The most germane comparison is to the UVA system from Uijlings et al. [39], since our systems use the same region proposal algorithm. To classify regions, their method builds a four-level spatial pyramid and populates it with densely sampled SIFT, Extended OpponentSIFT, and RGB-SIFT descriptors, each vector quantized with 4000-word codebooks. Classification is performed with a histogram intersection kernel SVM. Compared to their multi-feature, non-linear kernel SVM approach, we achieve a large improvement in mAP, from 35.1% to 53.7% mAP, while also being much faster (Section 2.2). Our method achieves similar performance (53.3% mAP) on VOC 2011/12 test.

2.5. Results on ILSVRC2013 detection

We ran R-CNN on the 200-class ILSVRC2013 detection dataset using the same system hyperparameters that we used for PASCAL VOC. We followed the same protocol of submitting test results to the ILSVRC2013 evaluation server only twice, once with and once without bounding-box regression.

Figure 3 compares R-CNN to the entries in the ILSVRC 2013 competition and to the post-competition OverFeat result [34]. R-CNN achieves a mAP of 31.4%, which is significantly ahead of the second-best result of 24.3% from OverFeat. To give a sense of the AP distribution over classes, box plots are also presented and a table of per-class APs follows at the end of the paper in Table 8. Most of the competing submissions (OverFeat, NEC-MU, UvA-Euvision, Toronto A, and UIUC-IFP) used convolutional neural networks, indicating that there is significant nuance in how CNNs can be applied to object detection, leading to greatly varying outcomes.

In Section 4, we give an overview of the ILSVRC2013 detection dataset and provide details about choices that we made when running R-CNN on it.

3. Visualization, ablation, and modes of error

3.1. Visualizing learned features

First-layer filters can be visualized directly and are easy to understand [25]. They capture oriented edges and opponent colors. Understanding the subsequent layers is more challenging. Zeiler and Fergus present a visually attractive deconvolutional approach in [42]. We propose a simple (and complementary) non-parametric method that directly shows what the network learned.

The idea is to single out a particular unit (feature) in the network and use it as if it were an object detector in its own right. That is, we compute the unit's activations on a large set of held-out region proposals (about 10 million), sort the proposals from highest to lowest activation, perform non-maximum suppression, and then display the top-scoring regions. Our method lets the selected unit "speak for itself" by showing exactly which inputs it fires on. We avoid averaging in order to see different visual modes and gain insight into the invariances computed by the unit.

VOC 2010 test	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	sheep	sofa	train	tv	mAP
DPM v5 [20]†	49.2	53.8	13.1	15.3	35.5	53.4	49.7	27.0	17.2	28.8	14.7	17.8	46.4	51.2	47.7	10.8	34.2	20.7	43.8	38.3	33.4
UVA [39]	56.2	42.4	15.3	12.6	21.8	49.3	36.8	46.1	12.9	32.1	30.0	36.5	43.5	52.9	32.9	15.3	41.1	31.8	47.0	44.8	35.1
Regionlets [41]	65.0	48.9	25.9	24.6	24.5	56.1	54.5	51.2	17.0	28.9	30.2	35.8	40.2	55.7	43.5	14.3	43.9	32.6	54.0	45.9	39.7
SegDPM [18]†	61.4	53.4	25.6	25.2	35.5	51.7	50.6	50.8	19.3	33.8	26.8	40.4	48.3	54.4	47.1	14.8	38.7	35.0	52.8	43.1	40.4
R-CNN	67.1	64.1	46.7	32.0	30.5	56.4	57.2	65.9	27.0	47.3	40.9	66.6	57.8	65.9	53.6	26.7	56.5	38.1	52.8	50.2	50.2
R-CNN BB	**71.8**	**65.8**	**53.0**	**36.8**	**35.9**	**59.7**	**60.0**	**69.9**	**27.9**	**50.6**	**41.4**	**70.0**	**62.0**	**69.0**	**58.1**	**29.5**	**59.4**	**39.3**	**61.2**	**52.4**	**53.7**

Table 1: Detection average precision (%) on VOC 2010 test. R-CNN is most directly comparable to UVA and Regionlets since all methods use selective search region proposals. Bounding-box regression (BB) is described in Section C. At publication time, SegDPM was the top-performer on the PASCAL VOC leaderboard. †DPM and SegDPM use context rescoring not used by the other methods.

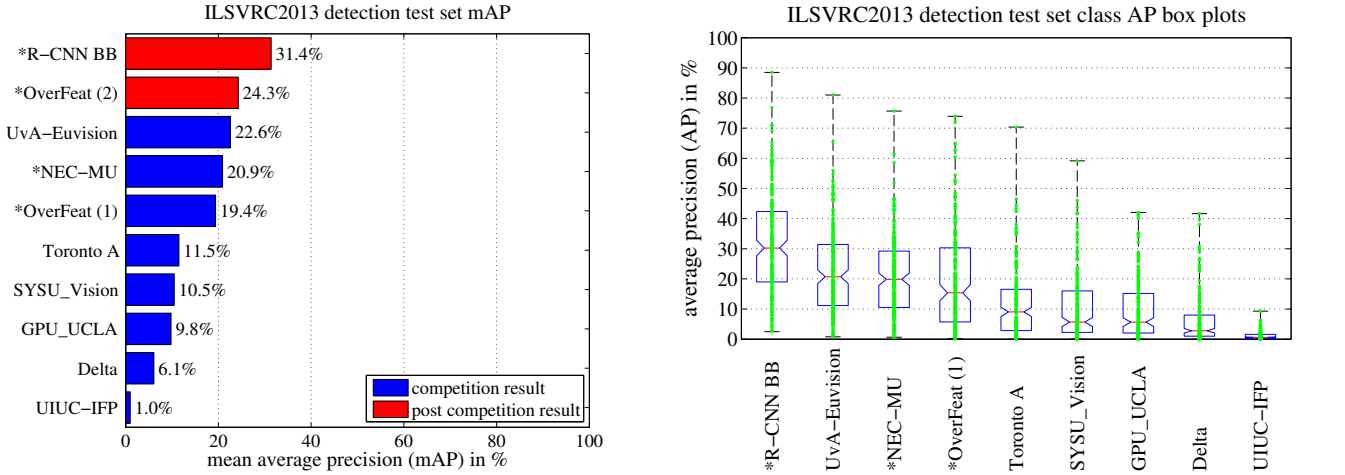


Figure 3: (Left) Mean average precision on the ILSVRC2013 detection test set. Methods preceeded by * use outside training data (images and labels from the ILSVRC classification dataset in all cases). **(Right) Box plots for the 200 average precision values per method.** A box plot for the post-competition OverFeat result is not shown because per-class APs are not yet available (per-class APs for R-CNN are in Table 8 and also included in the tech report source uploaded to arXiv.org; see `R-CNN-ILSVRC2013-APs.txt`). The red line marks the median AP, the box bottom and top are the 25th and 75th percentiles. The whiskers extend to the min and max AP of each method. Each AP is plotted as a green dot over the whiskers (best viewed digitally with zoom).

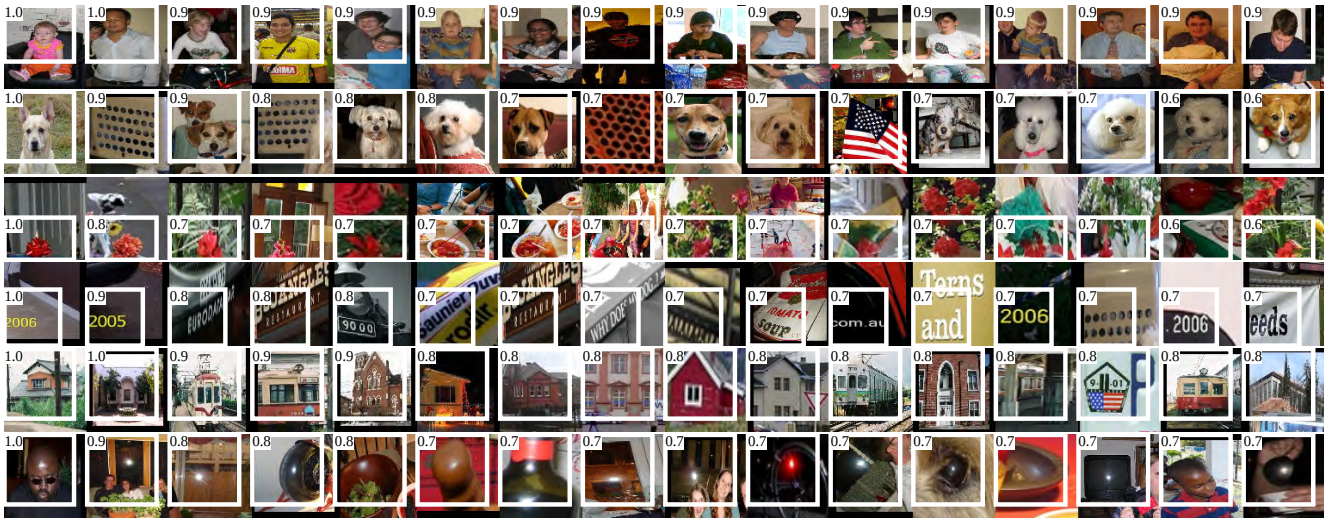


Figure 4: Top regions for six pool$_5$ units. Receptive fields and activation values are drawn in white. Some units are aligned to concepts, such as people (row 1) or text (4). Other units capture texture and material properties, such as dot arrays (2) and specular reflections (6).

VOC 2007 test	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	sheep	sofa	train	tv	mAP
R-CNN $pool_5$	51.8	60.2	36.4	27.8	23.2	52.8	60.6	49.2	18.3	47.8	44.3	40.8	56.6	58.7	42.4	23.4	46.1	36.7	51.3	55.7	44.2
R-CNN fc_6	59.3	61.8	43.1	34.0	25.1	53.1	60.6	52.8	21.7	47.8	42.7	47.8	52.5	58.5	44.6	25.6	48.3	34.0	53.1	58.0	46.2
R-CNN fc_7	57.6	57.9	38.5	31.8	23.7	51.2	58.9	51.4	20.0	50.5	40.9	46.0	51.6	55.9	43.3	23.3	48.1	35.3	51.0	57.4	44.7
R-CNN FT $pool_5$	58.2	63.3	37.9	27.6	26.1	54.1	66.9	51.4	26.7	55.5	43.4	43.1	57.7	59.0	45.8	28.1	50.8	40.6	53.1	56.4	47.3
R-CNN FT fc_6	63.5	66.0	47.9	37.7	29.9	62.5	70.2	60.2	32.0	57.9	47.0	53.5	60.1	64.2	52.2	31.3	55.0	50.0	57.7	63.0	53.1
R-CNN FT fc_7	64.2	69.7	50.0	41.9	32.0	62.6	71.0	60.7	32.7	58.5	46.5	56.1	60.6	66.8	54.2	31.5	52.8	48.9	57.9	64.7	54.2
R-CNN FT fc_7 BB	**68.1**	**72.8**	**56.8**	**43.0**	**36.8**	**66.3**	**74.2**	**67.6**	**34.4**	**63.5**	**54.5**	**61.2**	**69.1**	**68.6**	**58.7**	**33.4**	**62.9**	**51.1**	**62.5**	**64.8**	**58.5**
DPM v5 [20]	33.2	60.3	10.2	16.1	27.3	54.3	58.2	23.0	20.0	24.1	26.7	12.7	58.1	48.2	43.2	12.0	21.1	36.1	46.0	43.5	33.7
DPM ST [28]	23.8	58.2	10.5	8.5	27.1	50.4	52.0	7.3	19.2	22.8	18.1	8.0	55.9	44.8	32.4	13.3	15.9	22.8	46.2	44.9	29.1
DPM HSC [31]	32.2	58.3	11.5	16.3	30.6	49.9	54.8	23.5	21.5	27.7	34.0	13.7	58.1	51.6	39.9	12.4	23.5	34.4	47.4	45.2	34.3

Table 2: Detection average precision (%) on VOC 2007 test. Rows 1-3 show R-CNN performance without fine-tuning. Rows 4-6 show results for the CNN pre-trained on ILSVRC 2012 and then fine-tuned (FT) on VOC 2007 trainval. Row 7 includes a simple bounding-box regression (BB) stage that reduces localization errors (Section C). Rows 8-10 present DPM methods as a strong baseline. The first uses only HOG, while the next two use different feature learning approaches to augment or replace HOG.

VOC 2007 test	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	sheep	sofa	train	tv	mAP
R-CNN T-Net	64.2	69.7	50.0	41.9	32.0	62.6	71.0	60.7	32.7	58.5	46.5	56.1	60.6	66.8	54.2	31.5	52.8	48.9	57.9	64.7	54.2
R-CNN T-Net BB	68.1	72.8	56.8	43.0	36.8	66.3	74.2	67.6	34.4	63.5	54.5	61.2	69.1	68.6	58.7	33.4	62.9	51.1	62.5	64.8	58.5
R-CNN O-Net	71.6	73.5	58.1	42.2	39.4	70.7	76.0	74.5	38.7	71.0	56.9	74.5	67.9	69.6	59.3	**35.7**	62.1	64.0	66.5	**71.2**	62.2
R-CNN O-Net BB	**73.4**	**77.0**	**63.4**	**45.4**	**44.6**	**75.1**	**78.1**	**79.8**	**40.5**	**73.7**	**62.2**	**79.4**	**78.1**	**73.1**	**64.2**	35.6	**66.8**	**67.2**	**70.4**	71.1	**66.0**

Table 3: Detection average precision (%) on VOC 2007 test for two different CNN architectures. The first two rows are results from Table 2 using Krizhevsky et al.'s architecture (T-Net). Rows three and four use the recently proposed 16-layer architecture from Simonyan and Zisserman (O-Net) [43].

We visualize units from layer $pool_5$, which is the max-pooled output of the network's fifth and final convolutional layer. The $pool_5$ feature map is $6 \times 6 \times 256 = 9216$-dimensional. Ignoring boundary effects, each $pool_5$ unit has a receptive field of 195×195 pixels in the original 227×227 pixel input. A central $pool_5$ unit has a nearly global view, while one near the edge has a smaller, clipped support.

Each row in Figure 4 displays the top 16 activations for a $pool_5$ unit from a CNN that we fine-tuned on VOC 2007 trainval. Six of the 256 functionally unique units are visualized (Appendix D includes more). These units were selected to show a representative sample of what the network learns. In the second row, we see a unit that fires on dog faces and dot arrays. The unit corresponding to the third row is a red blob detector. There are also detectors for human faces and more abstract patterns such as text and triangular structures with windows. The network appears to learn a representation that combines a small number of class-tuned features together with a distributed representation of shape, texture, color, and material properties. The subsequent fully connected layer fc_6 has the ability to model a large set of compositions of these rich features.

3.2. Ablation studies

Performance layer-by-layer, without fine-tuning. To understand which layers are critical for detection performance, we analyzed results on the VOC 2007 dataset for each of the CNN's last three layers. Layer $pool_5$ was briefly described in Section 3.1. The final two layers are summarized below.

Layer fc_6 is fully connected to $pool_5$. To compute features, it multiplies a 4096×9216 weight matrix by the $pool_5$ feature map (reshaped as a 9216-dimensional vector) and then adds a vector of biases. This intermediate vector is component-wise half-wave rectified ($x \leftarrow \max(0, x)$).

Layer fc_7 is the final layer of the network. It is implemented by multiplying the features computed by fc_6 by a 4096×4096 weight matrix, and similarly adding a vector of biases and applying half-wave rectification.

We start by looking at results from the CNN *without fine-tuning* on PASCAL, i.e. all CNN parameters were pre-trained on ILSVRC 2012 only. Analyzing performance layer-by-layer (Table 2 rows 1-3) reveals that features from fc_7 generalize worse than features from fc_6. This means that 29%, or about 16.8 million, of the CNN's parameters can be removed without degrading mAP. More surprising is that removing *both* fc_7 and fc_6 produces quite good results even though $pool_5$ features are computed using *only 6%* of the CNN's parameters. Much of the CNN's representational power comes from its convolutional layers, rather than from the much larger densely connected layers. This finding suggests potential utility in computing a dense feature map, in the sense of HOG, of an arbitrary-sized image by using only the convolutional layers of the CNN. This representation would enable experimentation with sliding-window detectors, including DPM, on top of $pool_5$ features.

Performance layer-by-layer, with fine-tuning. We now look at results from our CNN after having fine-tuned its pa-

rameters on VOC 2007 trainval. The improvement is striking (Table 2 rows 4-6): fine-tuning increases mAP by 8.0 percentage points to 54.2%. The boost from fine-tuning is much larger for fc_6 and fc_7 than for $pool_5$, which suggests that the $pool_5$ features learned from ImageNet are general and that most of the improvement is gained from learning domain-specific non-linear classifiers on top of them.

Comparison to recent feature learning methods. Relatively few feature learning methods have been tried on PASCAL VOC detection. We look at two recent approaches that build on deformable part models. For reference, we also include results for the standard HOG-based DPM [20].

The first DPM feature learning method, DPM ST [28], augments HOG features with histograms of "sketch token" probabilities. Intuitively, a sketch token is a tight distribution of contours passing through the center of an image patch. Sketch token probabilities are computed at each pixel by a random forest that was trained to classify 35×35 pixel patches into one of 150 sketch tokens or background.

The second method, DPM HSC [31], replaces HOG with histograms of sparse codes (HSC). To compute an HSC, sparse code activations are solved for at each pixel using a learned dictionary of 100 7×7 pixel (grayscale) atoms. The resulting activations are rectified in three ways (full and both half-waves), spatially pooled, unit ℓ_2 normalized, and then power transformed ($x \leftarrow \text{sign}(x)|x|^\alpha$).

All R-CNN variants strongly outperform the three DPM baselines (Table 2 rows 8-10), including the two that use feature learning. Compared to the latest version of DPM, which uses only HOG features, our mAP is more than 20 percentage points higher: 54.2% vs. 33.7%—*a 61% relative improvement*. The combination of HOG and sketch tokens yields 2.5 mAP points over HOG alone, while HSC improves over HOG by 4 mAP points (when compared internally to their private DPM baselines—both use non-public implementations of DPM that underperform the open source version [20]). These methods achieve mAPs of 29.1% and 34.3%, respectively.

3.3. Network architectures

Most results in this paper use the network architecture from Krizhevsky et al. [25]. However, we have found that the choice of architecture has a large effect on R-CNN detection performance. In Table 3 we show results on VOC 2007 test using the 16-layer deep network recently proposed by Simonyan and Zisserman [43]. This network was one of the top performers in the recent ILSVRC 2014 classification challenge. The network has a homogeneous structure consisting of 13 layers of 3×3 convolution kernels, with five max pooling layers interspersed, and topped with three fully-connected layers. We refer to this network as "O-Net" for OxfordNet and the baseline as "T-Net" for TorontoNet.

To use O-Net in R-CNN, we downloaded the publicly available pre-trained network weights for the `VGG_ILSVRC_16_layers` model from the Caffe Model Zoo.[1] We then fine-tuned the network using the same protocol as we used for T-Net. The only difference was to use smaller minibatches (24 examples) as required in order to fit within GPU memory. The results in Table 3 show that R-CNN with O-Net substantially outperforms R-CNN with T-Net, increasing mAP from 58.5% to 66.0%. However there is a considerable drawback in terms of compute time, with the forward pass of O-Net taking roughly 7 times longer than T-Net.

3.4. Detection error analysis

We applied the excellent detection analysis tool from Hoiem et al. [23] in order to reveal our method's error modes, understand how fine-tuning changes them, and to see how our error types compare with DPM. A full summary of the analysis tool is beyond the scope of this paper and we encourage readers to consult [23] to understand some finer details (such as "normalized AP"). Since the analysis is best absorbed in the context of the associated plots, we present the discussion within the captions of Figure 5 and Figure 6.

3.5. Bounding-box regression

Based on the error analysis, we implemented a simple method to reduce localization errors. Inspired by the bounding-box regression employed in DPM [17], we train a linear regression model to predict a new detection window given the $pool_5$ features for a selective search region proposal. Full details are given in Appendix C. Results in Table 1, Table 2, and Figure 5 show that this simple approach fixes a large number of mislocalized detections, boosting mAP by 3 to 4 points.

3.6. Qualitative results

Qualitative detection results on ILSVRC2013 are presented in Figure 8 and Figure 9 at the end of the paper. Each image was sampled randomly from the val_2 set and all detections from all detectors with a precision greater than 0.5 are shown. Note that these are not curated and give a realistic impression of the detectors in action. More qualitative results are presented in Figure 10 and Figure 11, but these have been curated. We selected each image because it contained interesting, surprising, or amusing results. Here, also, all detections at precision greater than 0.5 are shown.

4. The ILSVRC2013 detection dataset

In Section 2 we presented results on the ILSVRC2013 detection dataset. This dataset is less homogeneous than

[1] https://github.com/BVLC/caffe/wiki/Model-Zoo

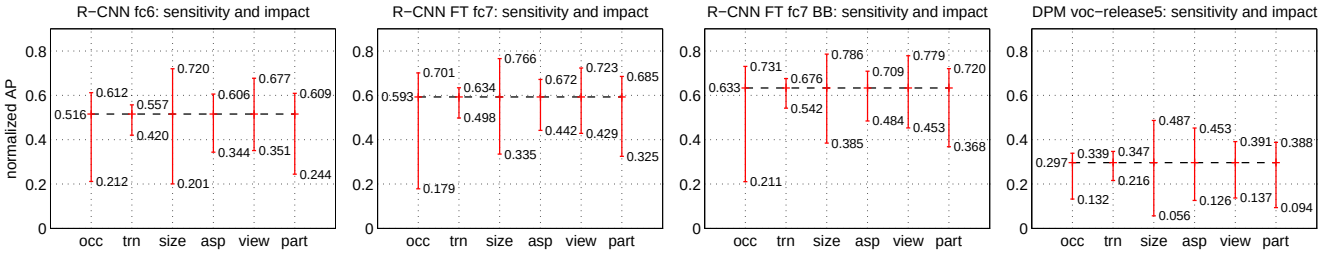


Figure 6: Sensitivity to object characteristics. Each plot shows the mean (over classes) normalized AP (see [23]) for the highest and lowest performing subsets within six different object characteristics (occlusion, truncation, bounding-box area, aspect ratio, viewpoint, part visibility). We show plots for our method (R-CNN) with and without fine-tuning (FT) and bounding-box regression (BB) as well as for DPM voc-release5. Overall, fine-tuning does not reduce sensitivity (the difference between max and min), but does substantially improve both the highest and lowest performing subsets for nearly all characteristics. This indicates that fine-tuning does more than simply improve the lowest performing subsets for aspect ratio and bounding-box area, as one might conjecture based on how we warp network inputs. Instead, fine-tuning improves robustness for all characteristics including occlusion, truncation, viewpoint, and part visibility.

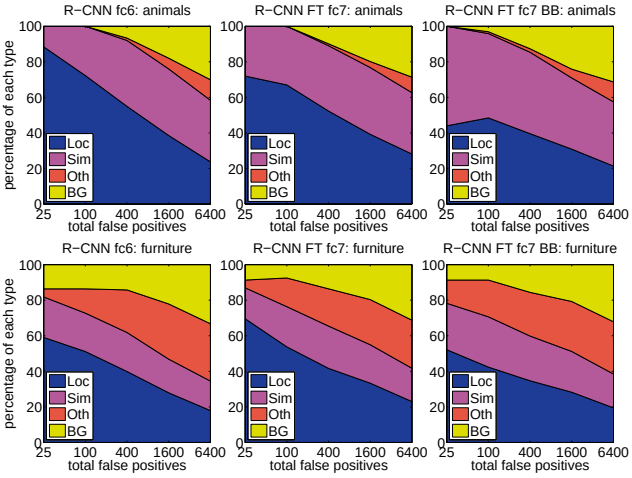


Figure 5: Distribution of top-ranked false positive (FP) types. Each plot shows the evolving distribution of FP types as more FPs are considered in order of decreasing score. Each FP is categorized into 1 of 4 types: Loc—poor localization (a detection with an IoU overlap with the correct class between 0.1 and 0.5, or a duplicate); Sim—confusion with a similar category; Oth—confusion with a dissimilar object category; BG—a FP that fired on background. Compared with DPM (see [23]), significantly more of our errors result from poor localization, rather than confusion with background or other object classes, indicating that the CNN features are much more discriminative than HOG. Loose localization likely results from our use of bottom-up region proposals and the positional invariance learned from pre-training the CNN for whole-image classification. Column three shows how our simple bounding-box regression method fixes many localization errors.

PASCAL VOC, requiring choices about how to use it. Since these decisions are non-trivial, we cover them in this section.

4.1. Dataset overview

The ILSVRC2013 detection dataset is split into three sets: train (395,918), val (20,121), and test (40,152), where the number of images in each set is in parentheses. The val and test splits are drawn from the same image distribution. These images are scene-like and similar in complexity (number of objects, amount of clutter, pose variability, etc.) to PASCAL VOC images. The val and test splits are exhaustively annotated, meaning that in each image all instances from all 200 classes are labeled with bounding boxes. The train set, in contrast, is drawn from the ILSVRC2013 *classification* image distribution. These images have more variable complexity with a skew towards images of a single centered object. Unlike val and test, the train images (due to their large number) are not exhaustively annotated. In any given train image, instances from the 200 classes may or may not be labeled. In addition to these image sets, each class has an extra set of negative images. Negative images are manually checked to validate that they do not contain any instances of their associated class. The negative image sets were not used in this work. More information on how ILSVRC was collected and annotated can be found in [11, 36].

The nature of these splits presents a number of choices for training R-CNN. The train images cannot be used for hard negative mining, because annotations are not exhaustive. Where should negative examples come from? Also, the train images have different statistics than val and test. Should the train images be used at all, and if so, to what extent? While we have not thoroughly evaluated a large number of choices, we present what seemed like the most obvious path based on previous experience.

Our general strategy is to rely heavily on the val set and use some of the train images as an auxiliary source of positive examples. To use val for both training and validation, we split it into roughly equally sized "val_1" and "val_2" sets. Since some classes have very few examples in val (the smallest has only 31 and half have fewer than 110), it is important to produce an approximately class-balanced partition. To do this, a large number of candidate splits were generated and the one with the smallest maximum relative

class imbalance was selected.[2] Each candidate split was generated by clustering val images using their class counts as features, followed by a randomized local search that may improve the split balance. The particular split used here has a maximum relative imbalance of about 11% and a median relative imbalance of 4%. The val_1/val_2 split and code used to produce them will be publicly available to allow other researchers to compare their methods on the val splits used in this report.

4.2. Region proposals

We followed the same region proposal approach that was used for detection on PASCAL. Selective search [39] was run in "fast mode" on each image in val_1, val_2, and test (but not on images in train). One minor modification was required to deal with the fact that selective search is not scale invariant and so the number of regions produced depends on the image resolution. ILSVRC image sizes range from very small to a few that are several mega-pixels, and so we resized each image to a fixed width (500 pixels) before running selective search. On val, selective search resulted in an average of 2403 region proposals per image with a 91.6% recall of all ground-truth bounding boxes (at 0.5 IoU threshold). This recall is notably lower than in PASCAL, where it is approximately 98%, indicating significant room for improvement in the region proposal stage.

4.3. Training data

For training data, we formed a set of images and boxes that includes all selective search and ground-truth boxes from val_1 together with up to N ground-truth boxes per class from train (if a class has fewer than N ground-truth boxes in train, then we take all of them). We'll call this dataset of images and boxes $\text{val}_1{+}\text{train}_N$. In an ablation study, we show mAP on val_2 for $N \in \{0, 500, 1000\}$ (Section 4.5).

Training data is required for three procedures in R-CNN: (1) CNN fine-tuning, (2) detector SVM training, and (3) bounding-box regressor training. CNN fine-tuning was run for 50k SGD iteration on $\text{val}_1{+}\text{train}_N$ using the exact same settings as were used for PASCAL. Fine-tuning on a single NVIDIA Tesla K20 took 13 hours using Caffe. For SVM training, all ground-truth boxes from $\text{val}_1{+}\text{train}_N$ were used as positive examples for their respective classes. Hard negative mining was performed on a randomly selected subset of 5000 images from val_1. An initial experiment indicated that mining negatives from all of val_1, versus a 5000 image subset (roughly half of it), resulted in only a 0.5 percentage point drop in mAP, while cutting SVM training time in half. No negative examples were taken from train because the annotations are not exhaustive. The extra sets of verified negative images were not used. The bounding-box regressors were trained on val_1.

[2] Relative imbalance is measured as $|a - b|/(a + b)$ where a and b are class counts in each half of the split.

4.4. Validation and evaluation

Before submitting results to the evaluation server, we validated data usage choices and the effect of fine-tuning and bounding-box regression on the val_2 set using the training data described above. All system hyperparameters (e.g., SVM C hyperparameters, padding used in region warping, NMS thresholds, bounding-box regression hyperparameters) were fixed at the same values used for PASCAL. Undoubtedly some of these hyperparameter choices are slightly suboptimal for ILSVRC, however the goal of this work was to produce a preliminary R-CNN result on ILSVRC without extensive dataset tuning. After selecting the best choices on val_2, we submitted exactly two result files to the ILSVRC2013 evaluation server. The first submission was without bounding-box regression and the second submission was with bounding-box regression. For these submissions, we expanded the SVM and bounding-box regressor training sets to use $\text{val}{+}\text{train}_{1\text{k}}$ and val, respectively. We used the CNN that was fine-tuned on $\text{val}_1{+}\text{train}_{1\text{k}}$ to avoid re-running fine-tuning and feature computation.

4.5. Ablation study

Table 4 shows an ablation study of the effects of different amounts of training data, fine-tuning, and bounding-box regression. A first observation is that mAP on val_2 matches mAP on test very closely. This gives us confidence that mAP on val_2 is a good indicator of test set performance. The first result, 20.9%, is what R-CNN achieves using a CNN pre-trained on the ILSVRC2012 classification dataset (no fine-tuning) and given access to the small amount of training data in val_1 (recall that half of the classes in val_1 have between 15 and 55 examples). Expanding the training set to $\text{val}_1{+}\text{train}_N$ improves performance to 24.1%, with essentially no difference between $N = 500$ and $N = 1000$. Fine-tuning the CNN using examples from just val_1 gives a modest improvement to 26.5%, however there is likely significant overfitting due to the small number of positive training examples. Expanding the fine-tuning set to $\text{val}_1{+}\text{train}_{1\text{k}}$, which adds up to 1000 positive examples per class from the train set, helps significantly, boosting mAP to 29.7%. Bounding-box regression improves results to 31.0%, which is a smaller relative gain that what was observed in PASCAL.

4.6. Relationship to OverFeat

There is an interesting relationship between R-CNN and OverFeat: OverFeat can be seen (roughly) as a special case of R-CNN. If one were to replace selective search region

test set	val_2	val_2	val_2	val_2	val_2	val_2	test	test
SVM training set	val_1	val_1+$train_{.5k}$	val_1+$train_{1k}$	val_1+$train_{1k}$	val_1+$train_{1k}$	val_1+$train_{1k}$	val+$train_{1k}$	val+$train_{1k}$
CNN fine-tuning set	n/a	n/a	n/a	val_1	val_1+$train_{1k}$	val_1+$train_{1k}$	val_1+$train_{1k}$	val_1+$train_{1k}$
bbox reg set	n/a	n/a	n/a	n/a	n/a	val_1	n/a	val
CNN feature layer	fc_6	fc_6	fc_6	fc_7	fc_7	fc_7	fc_7	fc_7
mAP	20.9	24.1	24.1	26.5	29.7	**31.0**	30.2	**31.4**
median AP	17.7	21.0	21.4	24.8	29.2	**29.6**	29.0	**30.3**

Table 4: ILSVRC2013 ablation study of data usage choices, fine-tuning, and bounding-box regression.

proposals with a multi-scale pyramid of regular square regions and change the per-class bounding-box regressors to a single bounding-box regressor, then the systems would be very similar (modulo some potentially significant differences in how they are trained: CNN detection fine-tuning, using SVMs, etc.). It is worth noting that OverFeat has a significant speed advantage over R-CNN: it is about 9x faster, based on a figure of 2 seconds per image quoted from [34]. This speed comes from the fact that OverFeat's sliding windows (i.e., region proposals) are not warped at the image level and therefore computation can be easily shared between overlapping windows. Sharing is implemented by running the entire network in a convolutional fashion over arbitrary-sized inputs. Speeding up R-CNN should be possible in a variety of ways and remains as future work.

5. Semantic segmentation

Region classification is a standard technique for semantic segmentation, allowing us to easily apply R-CNN to the PASCAL VOC segmentation challenge. To facilitate a direct comparison with the current leading semantic segmentation system (called O_2P for "second-order pooling") [4], we work within their open source framework. O_2P uses CPMC to generate 150 region proposals per image and then predicts the quality of each region, for each class, using support vector regression (SVR). The high performance of their approach is due to the quality of the CPMC regions and the powerful second-order pooling of multiple feature types (enriched variants of SIFT and LBP). We also note that Farabet et al. [16] recently demonstrated good results on several dense scene labeling datasets (not including PASCAL) using a CNN as a multi-scale per-pixel classifier.

We follow [2, 4] and extend the PASCAL segmentation training set to include the extra annotations made available by Hariharan et al. [22]. Design decisions and hyperparameters were cross-validated on the VOC 2011 validation set. Final test results were evaluated only once.

CNN features for segmentation. We evaluate three strategies for computing features on CPMC regions, all of which begin by warping the rectangular window around the region to 227×227. The first strategy (*full*) ignores the region's shape and computes CNN features directly on the warped window, exactly as we did for detection. However, these features ignore the non-rectangular shape of the region. Two regions might have very similar bounding boxes while having very little overlap. Therefore, the second strategy (*fg*) computes CNN features only on a region's foreground mask. We replace the background with the mean input so that background regions are zero after mean subtraction. The third strategy (*full*+*fg*) simply concatenates the *full* and *fg* features; our experiments validate their complementarity.

	full R-CNN		*fg* R-CNN		*full*+*fg* R-CNN	
O_2P [4]	fc_6	fc_7	fc_6	fc_7	fc_6	fc_7
46.4	43.0	42.5	43.7	42.1	**47.9**	45.8

Table 5: Segmentation mean accuracy (%) on VOC 2011 validation. Column 1 presents O_2P; 2-7 use our CNN pre-trained on ILSVRC 2012.

Results on VOC 2011. Table 5 shows a summary of our results on the VOC 2011 validation set compared with O_2P. (See Appendix E for complete per-category results.) Within each feature computation strategy, layer fc_6 always outperforms fc_7 and the following discussion refers to the fc_6 features. The *fg* strategy slightly outperforms *full*, indicating that the masked region shape provides a stronger signal, matching our intuition. However, *full*+*fg* achieves an average accuracy of 47.9%, our best result by a margin of 4.2% (also modestly outperforming O_2P), indicating that the context provided by the *full* features is highly informative even given the *fg* features. Notably, training the 20 SVRs on our *full*+*fg* features takes an hour on a single core, compared to 10+ hours for training on O_2P features.

In Table 6 we present results on the VOC 2011 test set, comparing our best-performing method, fc_6 (*full*+*fg*), against two strong baselines. Our method achieves the highest segmentation accuracy for 11 out of 21 categories, and the highest overall segmentation accuracy of 47.9%, averaged across categories (but likely ties with the O_2P result under any reasonable margin of error). Still better performance could likely be achieved by fine-tuning.

VOC 2011 test	bg	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	sheep	sofa	train	tv	mean
R&P [2]	83.4	46.8	18.9	36.6	31.2	42.7	57.3	47.4	44.1	8.1	39.4	**36.1**	36.3	49.5	48.3	50.7	26.3	47.2	22.1	42.0	43.2	40.8
O_2P [4]	**85.4**	**69.7**	22.3	45.2	**44.4**	46.9	66.7	57.8	56.2	**13.5**	**46.1**	32.3	41.2	**59.1**	55.3	51.0	**36.2**	50.4	**27.8**	46.9	**44.6**	47.6
ours (*full+fg* R-CNN fc_6)	84.2	66.9	**23.7**	**58.3**	37.4	**55.4**	**73.3**	**58.7**	**56.5**	9.7	45.5	29.5	**49.3**	40.1	**57.8**	**53.9**	33.8	**60.7**	22.7	**47.1**	41.3	**47.9**

Table 6: Segmentation accuracy (%) on VOC 2011 test. We compare against two strong baselines: the "Regions and Parts" (R&P) method of [2] and the second-order pooling (O_2P) method of [4]. Without any fine-tuning, our CNN achieves top segmentation performance, outperforming R&P and roughly matching O_2P.

6. Conclusion

In recent years, object detection performance had stagnated. The best performing systems were complex ensembles combining multiple low-level image features with high-level context from object detectors and scene classifiers. This paper presents a simple and scalable object detection algorithm that gives a 30% relative improvement over the best previous results on PASCAL VOC 2012.

We achieved this performance through two insights. The first is to apply high-capacity convolutional neural networks to bottom-up region proposals in order to localize and segment objects. The second is a paradigm for training large CNNs when labeled training data is scarce. We show that it is highly effective to pre-train the network—*with supervision*—for a auxiliary task with abundant data (image classification) and then to fine-tune the network for the target task where data is scarce (detection). We conjecture that the "supervised pre-training/domain-specific fine-tuning" paradigm will be highly effective for a variety of data-scarce vision problems.

We conclude by noting that it is significant that we achieved these results by using a combination of classical tools from computer vision *and* deep learning (bottom-up region proposals and convolutional neural networks). Rather than opposing lines of scientific inquiry, the two are natural and inevitable partners.

Acknowledgments. This research was supported in part by DARPA Mind's Eye and MSEE programs, by NSF awards IIS-0905647, IIS-1134072, and IIS-1212798, MURI N000014-10-1-0933, and by support from Toyota. The GPUs used in this research were generously donated by the NVIDIA Corporation.

Appendix

A. Object proposal transformations

The convolutional neural network used in this work requires a fixed-size input of 227×227 pixels. For detection, we consider object proposals that are arbitrary image rectangles. We evaluated two approaches for transforming object proposals into valid CNN inputs.

The first method ("tightest square with context") encloses each object proposal inside the tightest square and then scales (isotropically) the image contained in that square to the CNN input size. Figure 7 column (B) shows this transformation. A variant on this method ("tightest square without context") excludes the image content that surrounds the original object proposal. Figure 7 column (C) shows this transformation. The second method ("warp") anisotropically scales each object proposal to the CNN input size. Figure 7 column (D) shows the warp transformation.

Figure 7: Different object proposal transformations. (A) the original object proposal at its actual scale relative to the transformed CNN inputs; (B) tightest square with context; (C) tightest square without context; (D) warp. Within each column and example proposal, the top row corresponds to $p = 0$ pixels of context padding while the bottom row has $p = 16$ pixels of context padding.

For each of these transformations, we also consider including additional image context around the original object proposal. The amount of context padding (p) is defined as a border size around the original object proposal in the transformed input coordinate frame. Figure 7 shows $p = 0$ pixels in the top row of each example and $p = 16$ pixels in the bottom row. In all methods, if the source rectangle extends beyond the image, the missing data is replaced with the image mean (which is then subtracted before inputing the image into the CNN). A pilot set of experiments showed that warping with context padding ($p = 16$ pixels) outperformed the alternatives by a large margin (3-5 mAP points). Obviously more alternatives are possible, including using replication instead of mean padding. Exhaustive evaluation of these alternatives is left as future work.

B. Positive vs. negative examples and softmax

Two design choices warrant further discussion. The first is: Why are positive and negative examples defined differently for fine-tuning the CNN versus training the object detection SVMs? To review the definitions briefly, for fine-tuning we map each object proposal to the ground-truth instance with which it has maximum IoU overlap (if any) and label it as a positive for the matched ground-truth class if the IoU is at least 0.5. All other proposals are labeled "background" (i.e., negative examples for all classes). For training SVMs, in contrast, we take only the ground-truth boxes as positive examples for their respective classes and label proposals with less than 0.3 IoU overlap with all instances of a class as a negative for that class. Proposals that fall into the grey zone (more than 0.3 IoU overlap, but are not ground truth) are ignored.

Historically speaking, we arrived at these definitions because we started by training SVMs on features computed by the ImageNet pre-trained CNN, and so fine-tuning was not a consideration at that point in time. In that setup, we found that our particular label definition for training SVMs was optimal within the set of options we evaluated (which included the setting we now use for fine-tuning). When we started using fine-tuning, we initially used the same positive and negative example definition as we were using for SVM training. However, we found that results were much worse than those obtained using our current definition of positives and negatives.

Our hypothesis is that this difference in how positives and negatives are defined is not fundamentally important and arises from the fact that fine-tuning data is limited. Our current scheme introduces many "jittered" examples (those proposals with overlap between 0.5 and 1, but not ground truth), which expands the number of positive examples by approximately 30x. We conjecture that this large set is needed when fine-tuning the *entire* network to avoid overfitting. However, we also note that using these jittered examples is likely suboptimal because the network is not being fine-tuned for precise localization.

This leads to the second issue: Why, after fine-tuning, train SVMs at all? It would be cleaner to simply apply the last layer of the fine-tuned network, which is a 21-way softmax regression classifier, as the object detector. We tried this and found that performance on VOC 2007 dropped from 54.2% to 50.9% mAP. This performance drop likely arises from a combination of several factors including that the definition of positive examples used in fine-tuning does not emphasize precise localization and the softmax classifier was trained on randomly sampled negative examples rather than on the subset of "hard negatives" used for SVM training.

This result shows that it's possible to obtain close to the same level of performance without training SVMs after fine-tuning. We conjecture that with some additional tweaks to fine-tuning the remaining performance gap may be closed. If true, this would simplify and speed up R-CNN training with no loss in detection performance.

C. Bounding-box regression

We use a simple bounding-box regression stage to improve localization performance. After scoring each selective search proposal with a class-specific detection SVM, we predict a new bounding box for the detection using a class-specific bounding-box regressor. This is similar in spirit to the bounding-box regression used in deformable part models [17]. The primary difference between the two approaches is that here we regress from features computed by the CNN, rather than from geometric features computed on the inferred DPM part locations.

The input to our training algorithm is a set of N training pairs $\{(P^i, G^i)\}_{i=1,\ldots,N}$, where $P^i = (P^i_x, P^i_y, P^i_w, P^i_h)$ specifies the pixel coordinates of the center of proposal P^i's bounding box together with P^i's width and height in pixels. Hence forth, we drop the superscript i unless it is needed. Each ground-truth bounding box G is specified in the same way: $G = (G_x, G_y, G_w, G_h)$. Our goal is to learn a transformation that maps a proposed box P to a ground-truth box G.

We parameterize the transformation in terms of four functions $d_x(P)$, $d_y(P)$, $d_w(P)$, and $d_h(P)$. The first two specify a scale-invariant translation of the center of P's bounding box, while the second two specify log-space translations of the width and height of P's bounding box. After learning these functions, we can transform an input proposal P into a predicted ground-truth box $\hat{G}$ by applying the transformation

$$\hat{G}_x = P_w d_x(P) + P_x \tag{1}$$

$$\hat{G}_y = P_h d_y(P) + P_y \tag{2}$$

$$\hat{G}_w = P_w \exp(d_w(P)) \tag{3}$$

$$\hat{G}_h = P_h \exp(d_h(P)). \tag{4}$$

Each function $d_\star(P)$ (where $\star$ is one of x, y, h, w) is modeled as a linear function of the pool_5 features of proposal P, denoted by $\boldsymbol{\phi}_5(P)$. (The dependence of $\boldsymbol{\phi}_5(P)$ on the image data is implicitly assumed.) Thus we have $d_\star(P) = \mathbf{w}_\star^{\mathrm{T}} \boldsymbol{\phi}_5(P)$, where $\mathbf{w}_\star$ is a vector of learnable model parameters. We learn $\mathbf{w}_\star$ by optimizing the regularized least squares objective (ridge regression):

$$\mathbf{w}_\star = \underset{\hat{\mathbf{w}}_\star}{\operatorname{argmin}} \sum_i^N (t^i_\star - \hat{\mathbf{w}}_\star^{\mathrm{T}} \boldsymbol{\phi}_5(P^i))^2 + \lambda \left\| \hat{\mathbf{w}}_\star \right\|^2 . \tag{5}$$

The regression targets $t_\star$ for the training pair (P, G) are defined as

$$t_x = (G_x - P_x)/P_w \quad (6)$$
$$t_y = (G_y - P_y)/P_h \quad (7)$$
$$t_w = \log(G_w/P_w) \quad (8)$$
$$t_h = \log(G_h/P_h). \quad (9)$$

As a standard regularized least squares problem, this can be solved efficiently in closed form.

We found two subtle issues while implementing bounding-box regression. The first is that regularization is important: we set $\lambda = 1000$ based on a validation set. The second issue is that care must be taken when selecting which training pairs (P, G) to use. Intuitively, if P is far from all ground-truth boxes, then the task of transforming P to a ground-truth box G does not make sense. Using examples like P would lead to a hopeless learning problem. Therefore, we only learn from a proposal P if it is *nearby* at least one ground-truth box. We implement "nearness" by assigning P to the ground-truth box G with which it has maximum IoU overlap (in case it overlaps more than one) if and only if the overlap is greater than a threshold (which we set to 0.6 using a validation set). All unassigned proposals are discarded. We do this once for each object class in order to learn a set of class-specific bounding-box regressors.

At test time, we score each proposal and predict its new detection window only once. In principle, we could iterate this procedure (i.e., re-score the newly predicted bounding box, and then predict a new bounding box from it, and so on). However, we found that iterating does not improve results.

D. Additional feature visualizations

Figure 12 shows additional visualizations for 20 pool_5 units. For each unit, we show the 24 region proposals that maximally activate that unit out of the full set of approximately 10 million regions in all of VOC 2007 test.

We label each unit by its (y, x, channel) position in the $6 \times 6 \times 256$ dimensional pool_5 feature map. Within each channel, the CNN computes exactly the same function of the input region, with the (y, x) position changing only the receptive field.

E. Per-category segmentation results

In Table 7 we show the per-category segmentation accuracy on VOC 2011 val for each of our six segmentation methods in addition to the O_2P method [4]. These results show which methods are strongest across each of the 20 PASCAL classes, plus the background class.

F. Analysis of cross-dataset redundancy

One concern when training on an auxiliary dataset is that there might be redundancy between it and the test set. Even though the tasks of object detection and whole-image classification are substantially different, making such cross-set redundancy much less worrisome, we still conducted a thorough investigation that quantifies the extent to which PASCAL test images are contained within the ILSVRC 2012 training and validation sets. Our findings may be useful to researchers who are interested in using ILSVRC 2012 as training data for the PASCAL image classification task.

We performed two checks for duplicate (and near-duplicate) images. The first test is based on exact matches of flickr image IDs, which are included in the VOC 2007 test annotations (these IDs are intentionally kept secret for subsequent PASCAL test sets). All PASCAL images, and about half of ILSVRC, were collected from flickr.com. This check turned up 31 matches out of 4952 (0.63%).

The second check uses GIST [30] descriptor matching, which was shown in [13] to have excellent performance at near-duplicate image detection in large (> 1 million) image collections. Following [13], we computed GIST descriptors on warped 32×32 pixel versions of all ILSVRC 2012 trainval and PASCAL 2007 test images.

Euclidean distance nearest-neighbor matching of GIST descriptors revealed 38 near-duplicate images (including all 31 found by flickr ID matching). The matches tend to vary slightly in JPEG compression level and resolution, and to a lesser extent cropping. These findings show that the overlap is small, less than 1%. For VOC 2012, because flickr IDs are not available, we used the GIST matching method only. Based on GIST matches, 1.5% of VOC 2012 test images are in ILSVRC 2012 trainval. The slightly higher rate for VOC 2012 is likely due to the fact that the two datasets were collected closer together in time than VOC 2007 and ILSVRC 2012 were.

G. Document changelog

This document tracks the progress of R-CNN. To help readers understand how it has changed over time, here's a brief changelog describing the revisions.

v1 Initial version.

v2 CVPR 2014 camera-ready revision. Includes substantial improvements in detection performance brought about by (1) starting fine-tuning from a higher learning rate (0.001 instead of 0.0001), (2) using context padding when preparing CNN inputs, and (3) bounding-box regression to fix localization errors.

v3 Results on the ILSVRC2013 detection dataset and comparison with OverFeat were integrated into several sections (primarily Section 2 and Section 4).

VOC 2011 val	bg	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	sheep	sofa	train	tv	mean
O_2P [4]	**84.0**	**69.0**	21.7	47.7	42.2	42.4	**64.7**	**65.8**	57.4	**12.9**	37.4	20.5	43.7	35.7	52.7	51.0	**35.8**	**51.0**	28.4	59.8	49.7	46.4
full R-CNN fc_6	81.3	56.2	23.9	42.9	40.7	38.8	59.2	56.5	53.2	11.4	34.6	16.7	48.1	37.0	51.4	46.0	31.5	44.0	24.3	53.7	51.1	43.0
full R-CNN fc_7	81.0	52.8	**25.1**	43.8	40.5	42.7	55.4	57.7	51.3	8.7	32.5	11.5	48.1	37.0	50.5	46.4	30.2	42.1	21.2	57.7	**56.0**	42.5
fg R-CNN fc_6	81.4	54.1	21.1	40.6	38.7	**53.6**	59.9	57.2	52.5	9.1	36.5	**23.6**	46.4	38.1	53.2	51.3	32.2	38.7	**29.0**	53.0	47.5	43.7
fg R-CNN fc_7	80.9	50.1	20.0	40.2	34.1	40.9	59.7	59.8	52.7	7.3	32.1	14.3	48.8	42.9	54.0	48.6	28.9	42.6	24.9	52.2	48.8	42.1
full+*fg* R-CNN fc_6	83.1	60.4	23.2	48.4	**47.3**	52.6	61.6	60.6	**59.1**	10.8	**45.8**	20.9	**57.7**	43.3	**57.4**	**52.9**	34.7	48.7	28.1	60.0	48.6	**47.9**
full+*fg* R-CNN fc_7	82.3	56.7	20.6	**49.9**	44.2	43.6	59.3	61.3	57.8	7.7	38.4	15.1	53.4	**43.7**	50.8	52.0	34.1	47.8	24.7	**60.1**	55.2	45.7

Table 7: Per-category segmentation accuracy (%) on the VOC 2011 validation set.

v4 The softmax vs. SVM results in Appendix B contained an error, which has been fixed. We thank Sergio Guadarrama for helping to identify this issue.

v5 Added results using the new 16-layer network architecture from Simonyan and Zisserman [43] to Section 3.3 and Table 3.

References

[1] B. Alexe, T. Deselaers, and V. Ferrari. Measuring the objectness of image windows. *TPAMI*, 2012. 2

[2] P. Arbeláez, B. Hariharan, C. Gu, S. Gupta, L. Bourdev, and J. Malik. Semantic segmentation using regions and parts. In *CVPR*, 2012. 10, 11

[3] P. Arbeláez, J. Pont-Tuset, J. Barron, F. Marques, and J. Malik. Multiscale combinatorial grouping. In *CVPR*, 2014. 3

[4] J. Carreira, R. Caseiro, J. Batista, and C. Sminchisescu. Semantic segmentation with second-order pooling. In *ECCV*, 2012. 4, 10, 11, 13, 14

[5] J. Carreira and C. Sminchisescu. CPMC: Automatic object segmentation using constrained parametric min-cuts. *TPAMI*, 2012. 2, 3

[6] D. Cireşan, A. Giusti, L. Gambardella, and J. Schmidhuber. Mitosis detection in breast cancer histology images with deep neural networks. In *MICCAI*, 2013. 3

[7] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *CVPR*, 2005. 1

[8] T. Dean, M. A. Ruzon, M. Segal, J. Shlens, S. Vijayanarasimhan, and J. Yagnik. Fast, accurate detection of 100,000 object classes on a single machine. In *CVPR*, 2013. 3

[9] J. Deng, A. Berg, S. Satheesh, H. Su, A. Khosla, and L. Fei-Fei. ImageNet Large Scale Visual Recognition Competition 2012 (ILSVRC2012). http://www.image-net.org/challenges/LSVRC/2012/. 1

[10] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A large-scale hierarchical image database. In *CVPR*, 2009. 1

[11] J. Deng, O. Russakovsky, J. Krause, M. Bernstein, A. C. Berg, and L. Fei-Fei. Scalable multi-label annotation. In *CHI*, 2014. 8

[12] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell. DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition. In *ICML*, 2014. 2

[13] M. Douze, H. Jégou, H. Sandhawalia, L. Amsaleg, and C. Schmid. Evaluation of gist descriptors for web-scale image search. In *Proc. of the ACM International Conference on Image and Video Retrieval*, 2009. 13

[14] I. Endres and D. Hoiem. Category independent object proposals. In *ECCV*, 2010. 3

[15] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The PASCAL Visual Object Classes (VOC) Challenge. *IJCV*, 2010. 1, 4

[16] C. Farabet, C. Couprie, L. Najman, and Y. LeCun. Learning hierarchical features for scene labeling. *TPAMI*, 2013. 10

[17] P. Felzenszwalb, R. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part based models. *TPAMI*, 2010. 2, 4, 7, 12

[18] S. Fidler, R. Mottaghi, A. Yuille, and R. Urtasun. Bottom-up segmentation for top-down detection. In *CVPR*, 2013. 4, 5

[19] K. Fukushima. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological cybernetics*, 36(4):193–202, 1980. 1

[20] R. Girshick, P. Felzenszwalb, and D. McAllester. Discriminatively trained deformable part models, release 5. http://www.cs.berkeley.edu/~rbg/latent-v5/. 2, 5, 6, 7

[21] C. Gu, J. J. Lim, P. Arbeláez, and J. Malik. Recognition using regions. In *CVPR*, 2009. 2

[22] B. Hariharan, P. Arbeláez, L. Bourdev, S. Maji, and J. Malik. Semantic contours from inverse detectors. In *ICCV*, 2011. 10

[23] D. Hoiem, Y. Chodpathumwan, and Q. Dai. Diagnosing error in object detectors. In *ECCV*. 2012. 2, 7, 8

[24] Y. Jia. Caffe: An open source convolutional architecture for fast feature embedding. http://caffe.berkeleyvision.org/, 2013. 3

[25] A. Krizhevsky, I. Sutskever, and G. Hinton. ImageNet classification with deep convolutional neural networks. In *NIPS*, 2012. 1, 3, 4, 7

[26] Y. LeCun, B. Boser, J. Denker, D. Henderson, R. Howard, W. Hubbard, and L. Jackel. Backpropagation applied to handwritten zip code recognition. *Neural Comp.*, 1989. 1

[27] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proc. of the IEEE*, 1998. 1

[28] J. J. Lim, C. L. Zitnick, and P. Dollár. Sketch tokens: A learned mid-level representation for contour and object detection. In *CVPR*, 2013. 6, 7

class	AP	class	AP	class	AP	class	AP	class	AP
accordion	50.8	centipede	30.4	hair spray	13.8	pencil box	11.4	snowplow	69.2
airplane	50.0	chain saw	14.1	hamburger	34.2	pencil sharpener	9.0	soap dispenser	16.8
ant	31.8	chair	19.5	hammer	9.9	perfume	32.8	soccer ball	43.7
antelope	53.8	chime	24.6	hamster	46.0	person	41.7	sofa	16.3
apple	30.9	cocktail shaker	46.2	harmonica	12.6	piano	20.5	spatula	6.8
armadillo	54.0	coffee maker	21.5	harp	50.4	pineapple	22.6	squirrel	31.3
artichoke	45.0	computer keyboard	39.6	hat with a wide brim	40.5	ping-pong ball	21.0	starfish	45.1
axe	11.8	computer mouse	21.2	head cabbage	17.4	pitcher	19.2	stethoscope	18.3
baby bed	42.0	corkscrew	24.2	helmet	33.4	pizza	43.7	stove	8.1
backpack	2.8	cream	29.9	hippopotamus	38.0	plastic bag	6.4	strainer	9.9
bagel	37.5	croquet ball	30.0	horizontal bar	7.0	plate rack	15.2	strawberry	26.8
balance beam	32.6	crutch	23.7	horse	41.7	pomegranate	32.0	stretcher	13.2
banana	21.9	cucumber	22.8	hotdog	28.7	popsicle	21.2	sunglasses	18.8
band aid	17.4	cup or mug	34.0	iPod	59.2	porcupine	37.2	swimming trunks	9.1
banjo	55.3	diaper	10.1	isopod	19.5	power drill	7.9	swine	45.3
baseball	41.8	digital clock	18.5	jellyfish	23.7	pretzel	24.8	syringe	5.7
basketball	65.3	dishwasher	19.9	koala bear	44.3	printer	21.3	table	21.7
bathing cap	37.2	dog	76.8	ladle	3.0	puck	14.1	tape player	21.4
beaker	11.3	domestic cat	44.1	ladybug	58.4	punching bag	29.4	tennis ball	59.1
bear	62.7	dragonfly	27.8	lamp	9.1	purse	8.0	tick	42.6
bee	52.9	drum	19.9	laptop	35.4	rabbit	71.0	tie	24.6
bell pepper	38.8	dumbbell	14.1	lemon	33.3	racket	16.2	tiger	61.8
bench	12.7	electric fan	35.0	lion	51.3	ray	41.1	toaster	29.2
bicycle	41.1	elephant	56.4	lipstick	23.1	red panda	61.1	traffic light	24.7
binder	6.2	face powder	22.1	lizard	38.9	refrigerator	14.0	train	60.8
bird	70.9	fig	44.5	lobster	32.4	remote control	41.6	trombone	13.8
bookshelf	19.3	filing cabinet	20.6	maillot	31.0	rubber eraser	2.5	trumpet	14.4
bow tie	38.8	flower pot	20.2	maraca	30.1	rugby ball	34.5	turtle	59.1
bow	9.0	flute	4.9	microphone	4.0	ruler	11.5	tv or monitor	41.7
bowl	26.7	fox	59.3	microwave	40.1	salt or pepper shaker	24.6	unicycle	27.2
brassiere	31.2	french horn	24.2	milk can	33.3	saxophone	40.8	vacuum	19.5
burrito	25.7	frog	64.1	miniskirt	14.9	scorpion	57.3	violin	13.7
bus	57.5	frying pan	21.5	monkey	49.6	screwdriver	10.6	volleyball	59.7
butterfly	88.5	giant panda	42.5	motorcycle	42.2	seal	20.9	waffle iron	24.0
camel	37.6	goldfish	28.6	mushroom	31.8	sheep	48.9	washer	39.8
can opener	28.9	golf ball	51.3	nail	4.5	ski	9.0	water bottle	8.1
car	44.5	golfcart	47.9	neck brace	31.6	skunk	57.9	watercraft	40.9
cart	48.0	guacamole	32.3	oboe	27.5	snail	36.2	whale	48.6
cattle	32.3	guitar	33.1	orange	38.8	snake	33.8	wine bottle	31.2
cello	28.9	hair dryer	13.0	otter	22.2	snowmobile	58.8	zebra	49.6

Table 8: Per-class average precision (%) on the ILSVRC2013 detection test set.

[29] D. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 2004. 1

[30] A. Oliva and A. Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *IJCV*, 2001. 13

[31] X. Ren and D. Ramanan. Histograms of sparse codes for

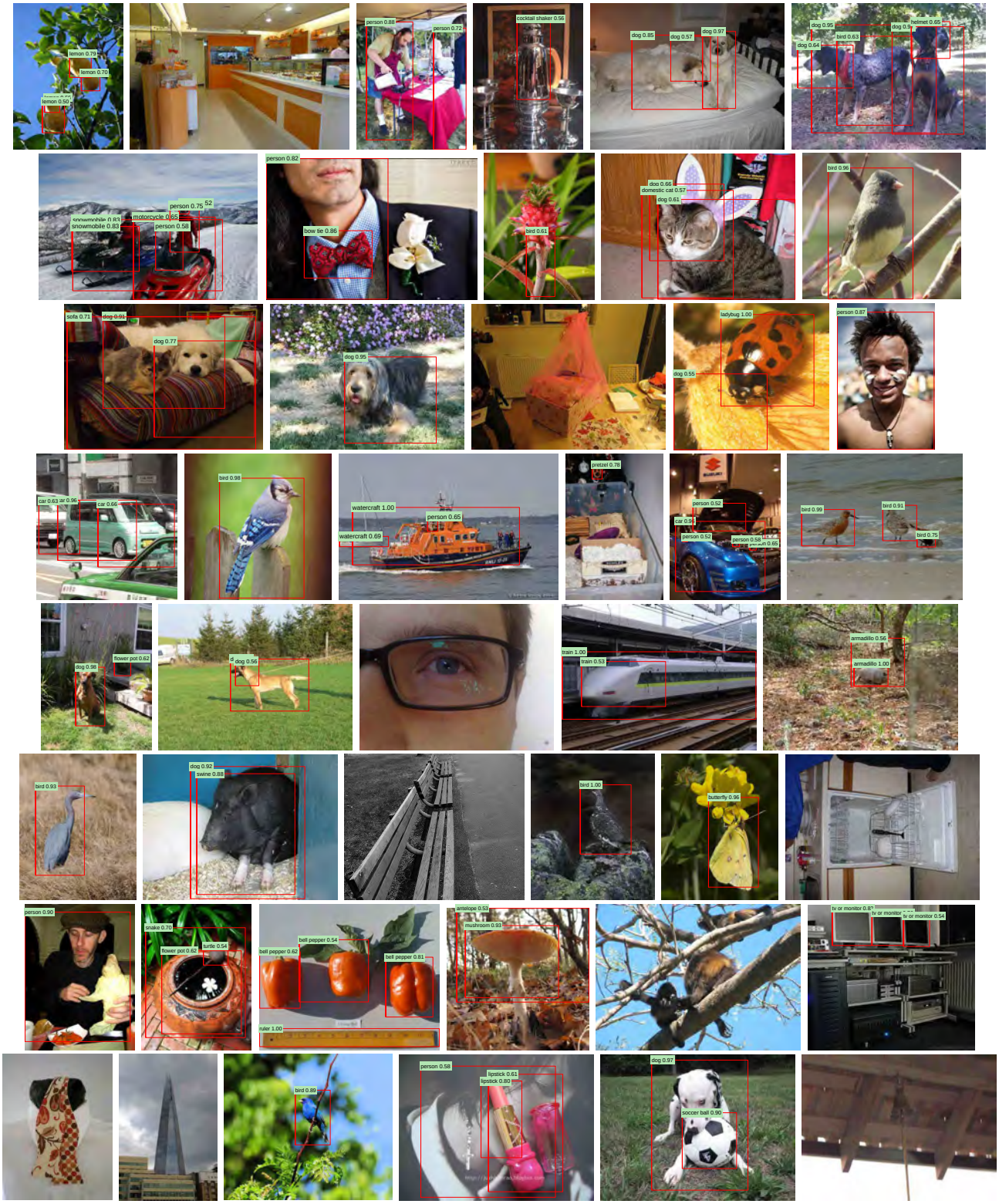


Figure 8: Example detections on the val_2 set from the configuration that achieved 31.0% mAP on val_2. Each image was sampled randomly (these are *not* curated). All detections at precision greater than 0.5 are shown. Each detection is labeled with the predicted class and the precision value of that detection from the detector's precision-recall curve. Viewing digitally with zoom is recommended.

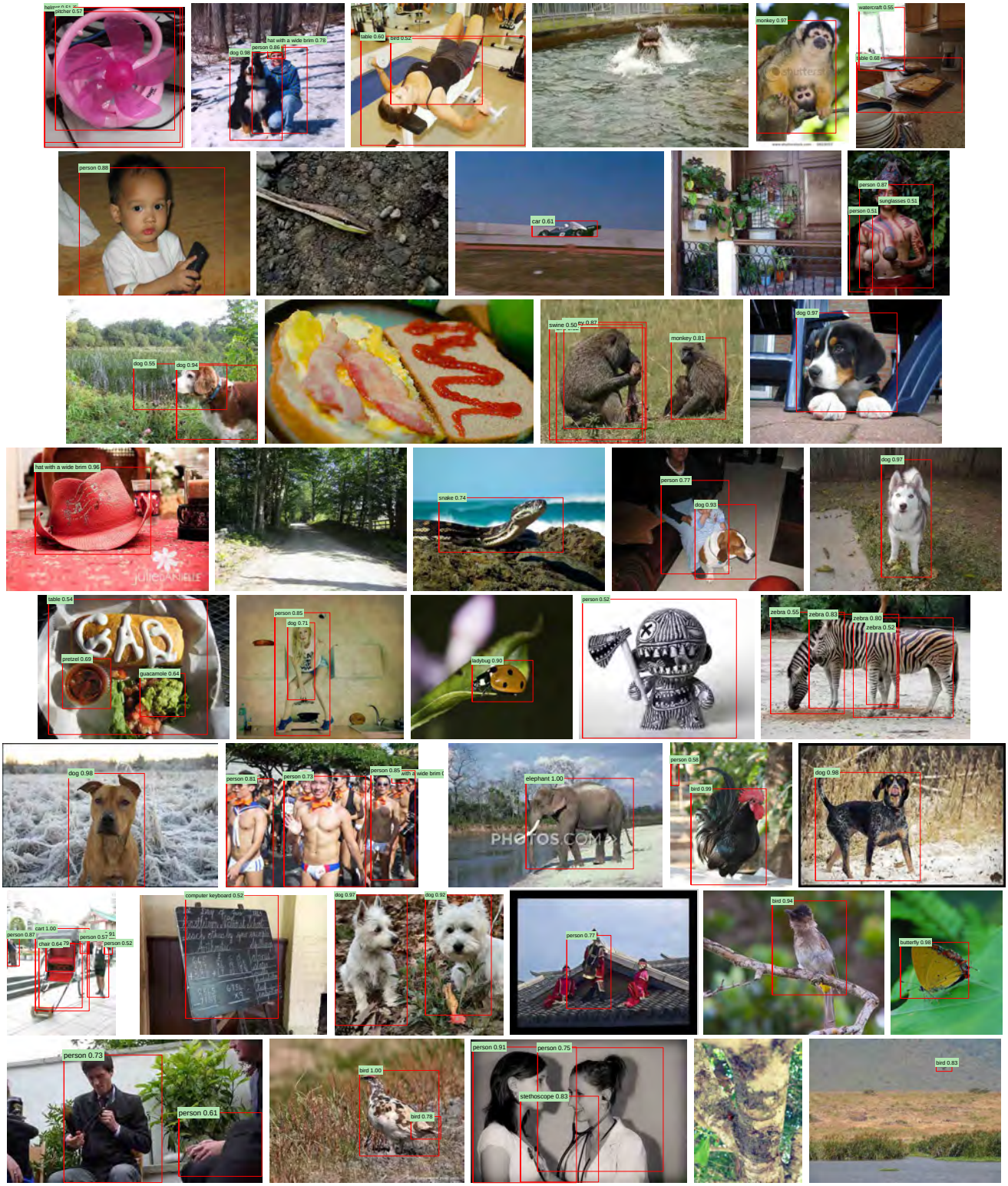


Figure 9: More randomly selected examples. See Figure 8 caption for details. Viewing digitally with zoom is recommended.

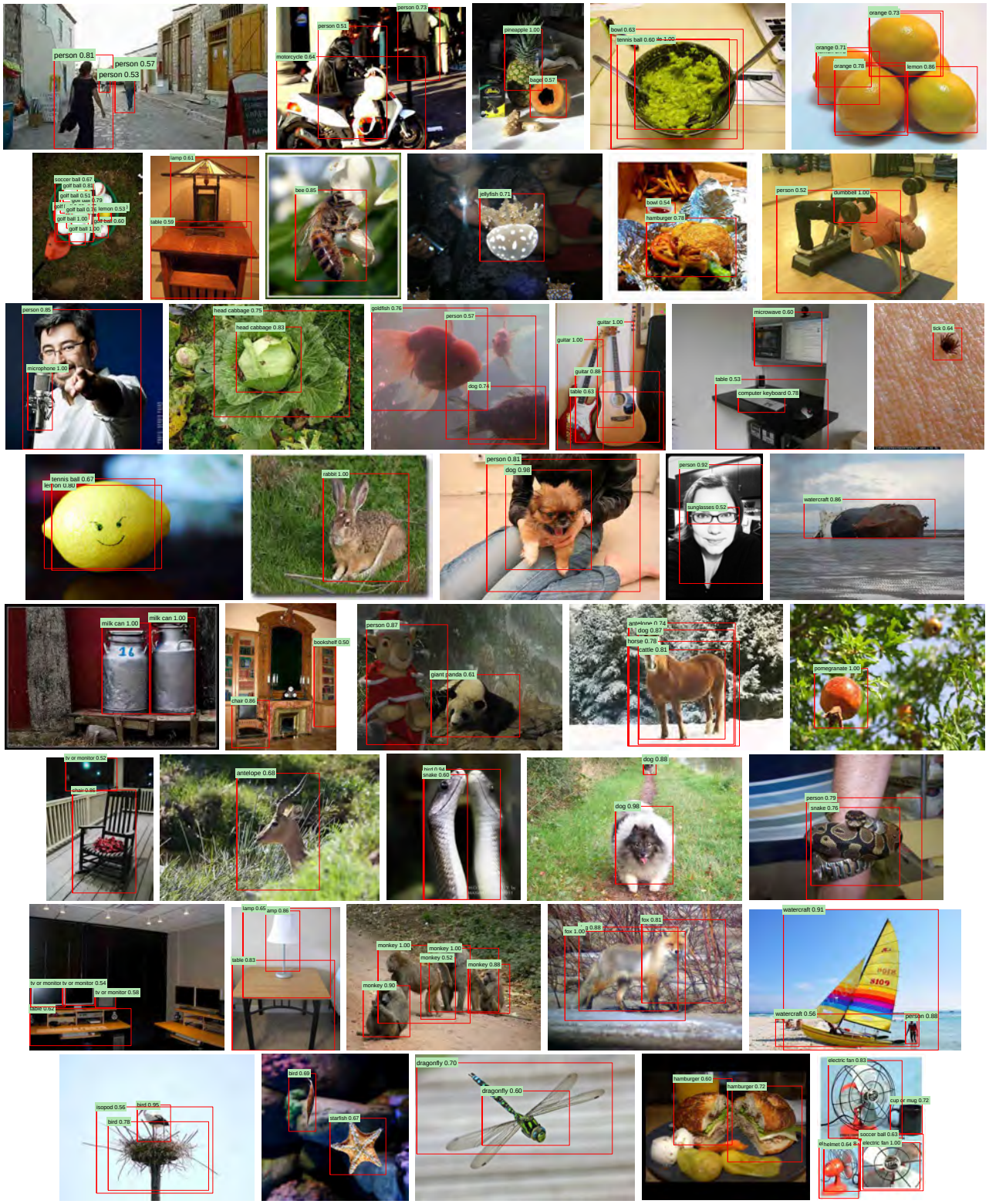


Figure 10: Curated examples. Each image was selected because we found it impressive, surprising, interesting, or amusing. Viewing digitally with zoom is recommended.

object detection. In *CVPR*, 2013. 6, 7

[32] H. A. Rowley, S. Baluja, and T. Kanade. Neural network-based face detection. *TPAMI*, 1998. 2

[33] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning internal representations by error propagation. *Parallel Distributed Processing*, 1:318–362, 1986. 1

[34] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun. OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks. In *ICLR*, 2014. 1, 2, 4, 10

[35] P. Sermanet, K. Kavukcuoglu, S. Chintala, and Y. LeCun. Pedestrian detection with unsupervised multi-stage feature learning. In *CVPR*, 2013. 2

[36] H. Su, J. Deng, and L. Fei-Fei. Crowdsourcing annotations for visual object detection. In *AAAI Technical Report, 4th Human Computation Workshop*, 2012. 8

[37] K. Sung and T. Poggio. Example-based learning for view-based human face detection. Technical Report A.I. Memo No. 1521, Massachussets Institute of Technology, 1994. 4

[38] C. Szegedy, A. Toshev, and D. Erhan. Deep neural networks for object detection. In *NIPS*, 2013. 2

[39] J. Uijlings, K. van de Sande, T. Gevers, and A. Smeulders. Selective search for object recognition. *IJCV*, 2013. 1, 2, 3, 4, 5, 9

[40] R. Vaillant, C. Monrocq, and Y. LeCun. Original approach for the localisation of objects in images. *IEE Proc on Vision, Image, and Signal Processing*, 1994. 2

[41] X. Wang, M. Yang, S. Zhu, and Y. Lin. Regionlets for generic object detection. In *ICCV*, 2013. 3, 5

[42] M. Zeiler, G. Taylor, and R. Fergus. Adaptive deconvolutional networks for mid and high level feature learning. In *CVPR*, 2011. 4

[43] K. Simonyan and A. Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv preprint*, arXiv:1409.1556, 2014. 6, 7, 14

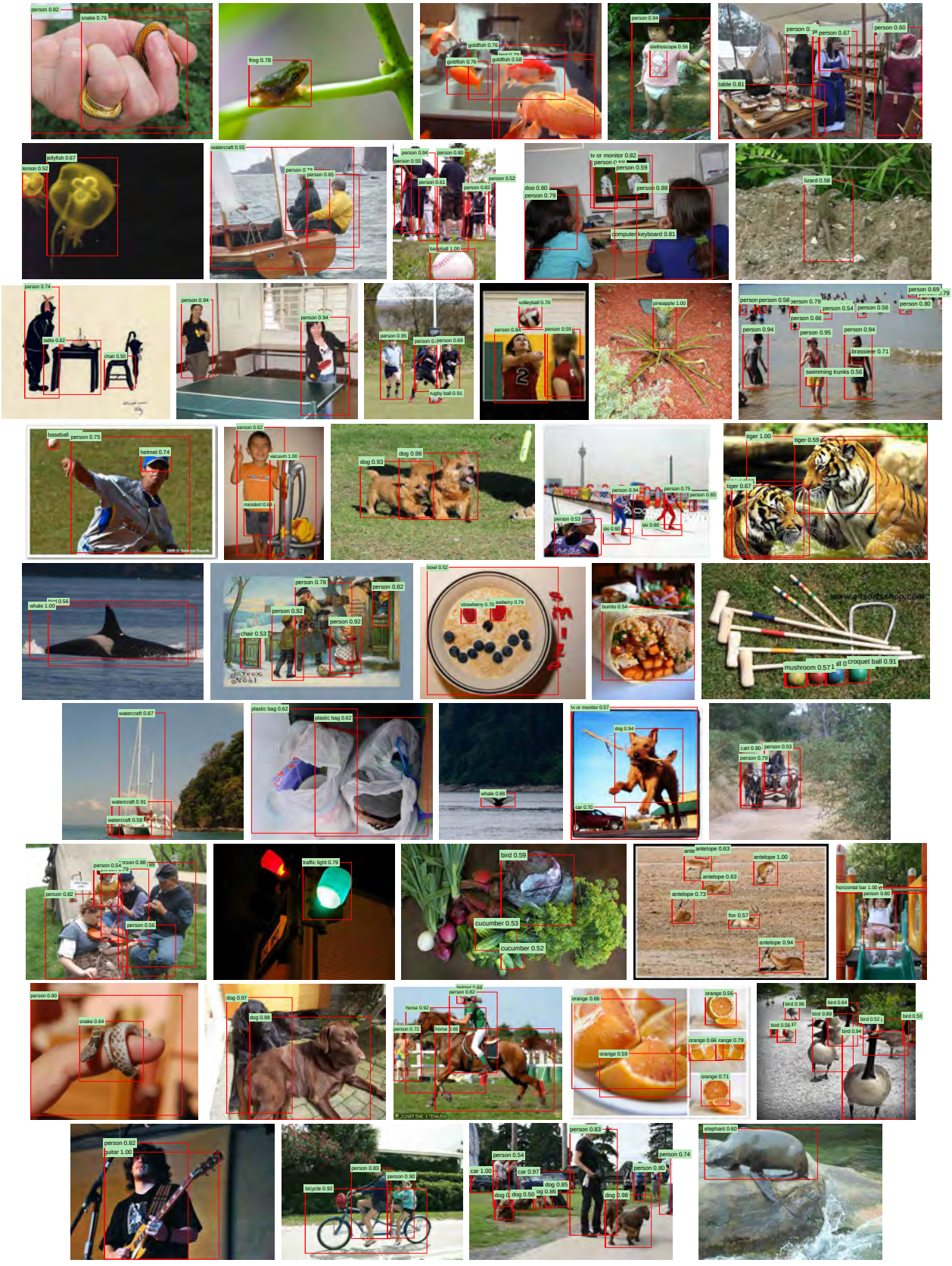


Figure 11: More curated examples. See Figure 10 caption for details. Viewing digitally with zoom is recommended.

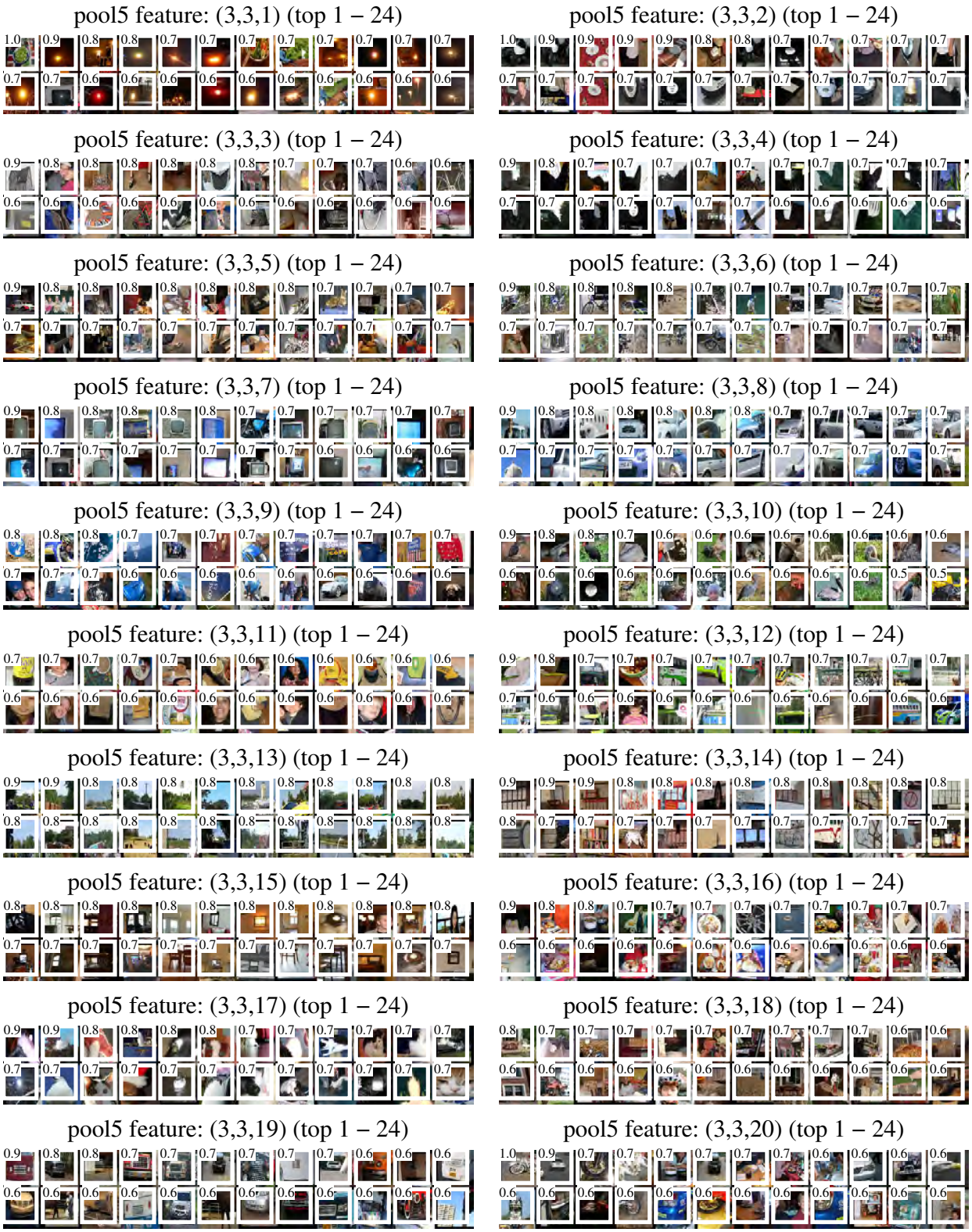


Figure 12: We show the 24 region proposals, out of the approximately 10 million regions in VOC 2007 test, that most strongly activate each of 20 units. Each montage is labeled by the unit's (y, x, channel) position in the $6 \times 6 \times 256$ dimensional pool_5 feature map. Each image region is drawn with an overlay of the unit's receptive field in white. The activation value (which we normalize by dividing by the max activation value over all units in a channel) is shown in the receptive field's upper-left corner. Best viewed digitally with zoom.

EXHIBIT R-4

A Good Practice Towards Top Performance of Face Recognition: Transferred Deep Feature Fusion

Lin Xiong[1*†], Jayashree Karlekar[1*], Jian Zhao[2*†], Jiashi Feng[2], *Member, IEEE*, Sugiri Pranata[1], and Shengmei Shen[1]

***Abstract*—Unconstrained face recognition performance evaluations have traditionally focused on Labeled Faces in the Wild (LFW) dataset for imagery and the YouTubeFaces (YTF) dataset for videos in the last couple of years. Spectacular progress in this field has resulted in a saturation on verification and identification accuracies for those benchmark datasets. In this paper, we propose a unified learning framework named transferred deep feature fusion targeting at the new IARPA Janus Bechmark A (IJB-A) face recognition dataset released by NIST face challenge. The IJB-A dataset includes real-world unconstrained faces from 500 subjects with full pose and illumination variations which are much harder than the LFW and YTF datasets. Inspired by transfer learning, we train two advanced deep convolutional neural networks (DCNN) with two different large datasets in source domain, respectively. By exploring the complementarity of two distinct DCNNs, deep feature fusion is utilized after feature extraction in target domain. Then, template specific linear SVMs is adopted to enhance the discrimination of framework. Finally, multiple matching scores corresponding different templates are merged as the final results. This simple unified framework outperforms the state-of-the-art by a wide margin on IJB-A dataset. Based on the proposed approach, we have submitted our IJB-A results to National Institute of Standards and Technology (NIST) for official evaluation.**

***Index Terms*—Face Recognition, Deep Convolutional Neural Network, Feature Fusion, Model Ensemble, SVMs.**

I. Introduction

FACE recognition performance using features of Deep Convolutional Neural Network (DCNN) have been dramatically improved in recent years. Many state-of-the-art algorithms claim very close [7],[12] or even have surpassed [13], [22],[27] human performance on Labeled Faces in the Wild (LFW) dataset. The saturation in recognition accuracy for current benchmark dataset has come. In order to push the development of frontier in regarding to unconstrained face recognition, a new face dataset template-based IJB-A is introduced recently [20], whose setting and solutions are aligned better with the requirements of real applications.

The IJB-A dataset is created to provide the latest and most challenging dataset for both verification and identification as shown is Fig.1. Unlike LFW and YTF, this dataset includes both image and video of subjects manually annotated with facial bounding boxes to avoid the near frontal condition, along with protocols for evaluation of both verification and identification. Those protocols significantly deviate from standard protocols for many face recognition algorithms [28],[29]. Moreover, the concept of template is introduced, simultaneously. A template refers to a collection of all media (images and/or video frames) of an interested face captured under different conditions that can be utilized as a combined single representation for matching task. The template-based setting reflects many real-world biometric scenarios, where capturing a subject's facial appearance is possible more than once under different acquisition ways. In other words, this new IJB-A face recognition task requires to deal with a more challenging set-to-set matching problem successfully regardless of face capture settings (illumination, sensor, resolution) or subject conditions (facial pose, expression, occlusion).

(a) Face recognition over single image.

(b) Unconstrained set-based face recognition.

Fig. 1: Comparison between face recognition over single image and unconstrained set-based face recognition. (a) Face recognition over single image. (b) Unconstrained set-based face recognition where each subject is represented by a set of mixed images and videos captured under unconstrained conditions. Each set contains large variations in face pose, expression, illumination and occlusion issues. Existing single-medium based recognition approaches cannot successfully address this problem consistently. Matched cases are bounded with green boxes, while non-matched cases are bounded with red boxes. Best viewed in color.

Our contributions can be summarized as following aspects:

1) A unified learning framework named transferred deep feature fusion is proposed for face verification and identification.
2) Two latest DCNN models are trained in source domain with two different large datasets in order to take full advantage of complementary between models and datasets.

[1]L. Xiong, J. Karlekar, S. Pranata and S.M. Shen are with Panasonic R&D Center Singapore, Singapore (lin.xiong@sg.panasonic.com; karlekar.jayashree@sg.panasonic.com; shengmei.shen@sg.panasonic.com).

[2]J. Zhao and J.S. Feng are with Department of Electrical and Computer Engineering, National University of Singapore, Singapore (zhaojian90@u.nus.edu; elefjia@nus.edu.sg). J. Zhao was an intern at Panasonic R&D Center Singapore during this work.

* L. Xiong, J. Zhao and J. Karlekar make an equal contribution.

† L. Xiong and J. Zhao are the corresponding author.

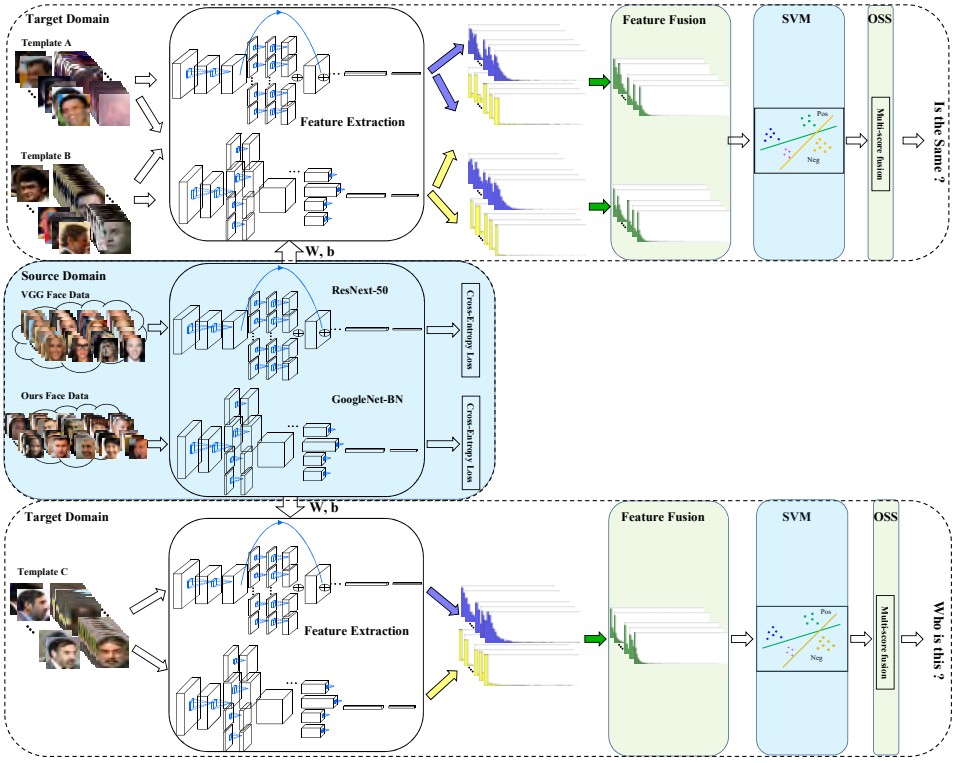


Fig. 2: Framework overview. Our learning framework consists three components: Deep feature learning module locates middle component, Template-based unconstrained face recognition is included in upper and lower components. Training procedures are illustrated with blue blocks, two-stage fusion is depicted in green blocks. Best viewed in color.

3) Two-stage fusion are designed, one for features and another for similarity scores.
4) One-vs-rest template specific linear SVMs with chosen negative set is trained in target domain.

In this paper, we propose a unified learning framework named transferred deep feature fusion. It can effectively integrate superiority of each module and outperform the state-of-the-art on IJB-A dataset. Inspired by transfer learning [1], facial feature encoding model of subjects are trained offline in a source domain, and this feature encoding model is transferred to a specific target domain where limited available faces of new subjects can be encoded. Specifically, in order to capture the intrinsic discrimination of subjects and enhance the generalization capability of face recognition models, we deploy two advanced deep convolutional neural networks (DCNN) with distinct architectures to learn the representation of faces on two different large datasets (each one has no overlap with IJB-A dataset) in source domain. These two DCNN models provide distinct feature representations which can better characterize the data distribution from different perspectives. The complementary between two distinct models is beneficial for feature representation [17]. Thus, representing a face from different perspectives could effectively decrease ambiguity among subjects and enhance the generalization performance of face recognition especially on extremely large number of subjects. After offline training procedure, those two DCNN models are transferred to target domain where templates of IJB-A dataset as inputs are performed feature extraction with shared weigths and biases, respectively. Then, features from two DCNN models are combined in order to obtain more discriminative representation. Finally, template specific linear SVMs are trained on fused features for classification. Furthermore, for set-to-set matching problem, multiple matching scores are merged into a single one [43],[45],[33] for each template pair as the final results. Comprehensive evaluations on IJB-A public dataset well demonstrate the significant superiority of the proposed learning framework over other state-of-the-art methods. Based on the proposed approach, we have submitted our IJB-A results to NIST for official evaluation.

This paper is organized as follows. We review the related work in Section II. Section III shows the details of transferred deep feature fusion. In Section IV, a comprehensive evaluation on IJB-A dataset is shown. Finally, the conclusion remarks and the future work are presented in Section V.

II. Related Work

Recently, all the top performing methods for face recognition on LFW and YTF are all based on DCNN architectures. Such as the VGG-Face model [14], as a typical application of the VGG-16 convolutional network architecture [8] trained on a reasonably and publicly large face dataset of 2.6M images of 2622 subjects, provides state-of-the-art performance. This dataset is called as VGG-Face data for convenience in the following section. FaceNet [22] utilizes the DCNN with inception module [18] for unconstrained face recognition. This network is trained using a private huge dataset of over 200M images and 8M subjects. DeepFace [7] deploys a DCNN coupled with 3D alignment, where facial pose is normalized by warping facial landmarks to a canonical position prior to encoding face images. DeepID2+ [12] and DeepID3 [13] extend the FaceNet model by including joint Bayesian metric learning [3] and multi-task learning. More better unconstrained face recognition performance is provided by them. Moreover, DeepFace is trained using a private dataset of 4.4M images and 4,030 subjects. DeepID2+ and DeepID3 are trained also using a private dataset of 202,595 images and 10,117 subjects with 25 networks and 50 networks, respectively. The idea of multiple model ensemble is involved. Moreover, many approaches use metric learning in the form of triplet loss similarity or joint Bayesian for the final loss to learn an optimal embedding for face recognition [22],[14],[27]. Thus, a recent study [16] concludes that multiple networks ensemble and metric learning are crucial for improvement on LFW. With

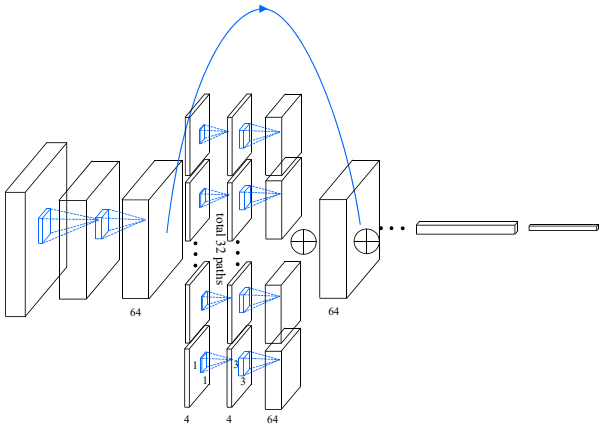


Fig. 3: A block of ResNext with cardinality=32.

the advent of IJB-A dataset introduced by NIST in 2015, the task of template-based unconstrained face recognition has attracted extensive attention. So far as we known, most algorithms for this challenging problem are also based on DCNN architecture as top performing methods did on LFW and YTF. Chen *et al.* [27] achieve good performance by extracting feature representations via a DCNN trained on public dataset which includes 490,356 images and 10,548 subjects. And then, those features as inputs are applied to learn metric matrix in order to project the feature vector into a low-dimensional space, meanwhile, maximizing the between-class variation and minimizing within-class variation via joint Bayesian metric learning. B-CNN [30] applies the bilinear CNN architecture to face identification. Deep Multi-pose [44] utilizes five pose specialized sub-networks with 3D pose rendering to encode multiple pose-specific features. Sensitivity of the recognition system to pose variations is reduced since an ensemble of pose-specific deep features is adopted. Pooling faces [45] aligns faces in 3D and bins them according to head pose and image quality. Pose-Aware Models (PAMs) [43] handles pose variability by learning Pose-Aware Models for frontal, half-profile and full-profile poses in order to improve face recognition performance in wild. Masi *et al.* [33] even question whether need to collect millions of faces or not for effective face recognition. Thus, a far more accessible means of increasing training data sizes is proposed. Pose, 3D shape and expression are utilized to synthesize more faces from CASIA-WebFace dataset [9]. Triplet Probabilistic Embedding (TPE) [42] couples a DCNN-based approach with a low-dimensional discriminative embedding learned using triplet probability constraints to solve the unconstrained face verification problem. TPE obtains better performance than previous algorithms on IJB-A dataset. Template Adaptation (TA) [34] proposes the idea of template adaptation which is a form of transfer learning to the set of media in a template. Combining DCNN features with template adaptation, it obtains better performance than TPE on IJB-A task. Ranjan *et al.* propose an all-in-one method [46] employed a multi-task learning framework that regularizes the shared parameters of CNN and builds a synergy among different domains and tasks. Until recently, Yang *et al.* propose Neural Aggregation Network (NAN) [47] which produces a compact and fixed-dimension feature representation. It adaptively aggregates the features to form a single feature inside the convex hull spanned by them. What's more interesting is that NAN learns to advocate high-quality face images while repelling low-quality ones such as blurred, occluded and improperly exposed faces. Thus, the face recognition performance on IJB-A dataset is pushed to reach an unprecedented height. Just a few days ago, Ranjan *et al.* [49] add an L_2-constraint to the feature descriptors which restricts them to lie on a hypersphere of a fixed radius. Therefore, minimizing the softmax loss is equivalent to maximizing the cosine similarity for the positive pairs and minimizing it for the negative pairs. In this way, the verification performance on IJB-A dataset is refreshed again.

In the current work, we also follow the similar way–DCNN model should be a good baseline. By virtue of the complementary between different DCNN architectures and datasets, we can obtain a more general feature representation model via ensemble strategy. Intrinsic discrimination of subjects is also important for face recognition, inspired by transfer learning, template specific linear one-vs-rest SVMs are trained in target domain. It shares similar idea as TA [34] while different negative set is chosen. Similar to [43],[45],[33], multiple matching scores are merged into a single one for set-to-set matching whereas an easier way is adopted. Last, we also deploy TPE to further enhance performance of face

recognition. More detailed information about our learning framework can be found in the next section part.

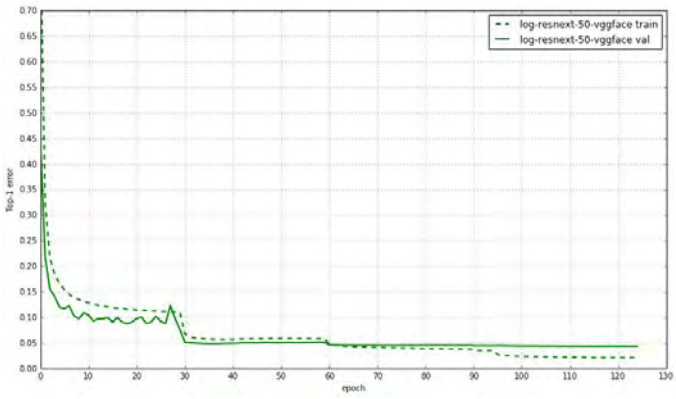


Fig. 4: Training on VGG-Face data. Solid curve denotes top 1 training error, and dotted line denotes validation error of the center crops.

III. Transferred Deep Feature Fusion

It is necessary that DCNN architectures are trained on tremendous dataset. However, IJB-A datasets contains 500 subjects with 5,396 images and 2,042 videos sampled to 20,412 frames in total. This is obviously inadequate. Unlike [33] where training data is increased by synthesizing faces based on pose, 3D shape and expression variations, inspired by domain adaptation, we need other huge labeled face datasets in source domain to train DCNN model. It is different from replacing the final entropy loss layer for a new task and fine-tuning the DCNN model on this new objective using data from the target domain [11]. We focus on training DCNN model and the one-vs-rest linear SVMs in source domain and target domain, separately. Last, one-shot-similarity (OSS) [2] is utilized to calculate similarity scores and we fuse those multiple matching scores into a single one for final performance evaluation. As shown in Fig.2, our learning framework consists three components: two distinct DCNN models are trained with two different large face datasets in source domain illustrated in middle component, respectively. In target domain, the new unseen data as inputs are fed into those two DCNN architectures with the shared weights and biases learned from source domain for feature extraction, respectively. Then, all features are combined in the first fusion stage. Template specific one-vs-rest SVMs are trained on those fused features in order to boost the intrinsic discrimination of subjects. Last but not least, multiple matching scores computed by OSS is weighted to one final score for verification and identification in the second fusion stage of upper and lower components, respectively. The detailed of each components of our learning framework are presented in the following subsections.

A. Deep feature learning in source domain

In this part, we discuss detailedly two DCNN models and two extra huge datasets for training in source domain.

Since Network-in-Network (NIN) [6] has been proposed, the depth of DCNN is refreshed again and again. Recent works [15],[40],[48] have shown that convolutional networks with small filters can be substantially deeper, more accurate, and efficient to train if they contain shorter connections between layers close to the input and those close to the output. The bypassing paths are presumed to be the key factor that eases the training of these very deep networks. This point is further supported by ResNets [32], in which pure identity mappings are used as bypassing paths. ResNets have achieved impressive, record-breaking performance on ImageNet [25]. Until recently, Xie *et al.* [39] reconstruct the building block of ResNets with aggregating a set of transformations. This simple design results in a homogeneous, multi-branch architecture that has only a few hyper-parameters to set. A new dimension called *cardinality* is proposed, which as an essential factor in addition to the dimension of depth and width. Thus, it is codenamed ResNext. A typical block of ResNext is shown in Fig.3. Considering the balance between performance and efficiency, we choose ResNext 50 as the first DCNN model.

For public large face dataset, the VGG-Face should be a better choice for ResNext 50. The original VGG-Face dataset includes 2,109,307 available images and 2,614 subjects. First, we utilize ground-truth bounding box given by dataset to crop and resize face images from the original ones. Each face image is 144×144. An off-the-shelf CNN model pre-trained on CASIA-WebFace is deployed to do noisy data cleaning. Moreover, the overlap subject with IJB-A dataset should be removed. Finally, we obtain 1,648,187 images and 2,613 subjects in total. For partition of training and validation parts, we refer to ImageNet. 90% of the total images (1,483,368) are served as training data. 5% of the total images (82,410) are viewed as validation data. Our implementation for VGG-Face on ResNext 50 is implemented by MXNet [26]. The image is resized from 144×144 to 480×480 for data augmentation. A 224×224 crop is randomly sampled from 480×480 or its horizontal flip, with the per-pixel mean substracted. The standard color augmentation [4] is used. We adopt batch normalization (BN) [19] right after each convolution and before ReLU. We initialize the weights as in [21] and train ResNext 50 from scratch. NAG with a mini-batch size of 256 is utilized on our GPU cluster machine. The learning rate starts from 0.1 and is divided by 10 every 30 epoch and the model is trained for up to 125 epoch. The weight decay is 0.0001 and the momentum is 0.9. The cardinality is 32. The training and validation curves are shown in Fig.4. Finally, we obtain the validation performance 95.63% at top1 and 97.00% at top 5, respectively.

Inspired by NIN, an orthogonal approach to making networks deeper (e.g., with the help of skip connections) is to increase the network width. The GoogLeNet [18] uses an "Inception module" which concatenates features maps produced by filers of different sizes. Different from ResNext which enhances representational power of network via extremely deep architecture, GoogLeNet depends on wider structure to boost capacity of network. Along with the BN emergence, training DCNN becomes easier than before. Thus, GoogLeNet-BN is our second DCNN model.

To train GoogLeNet-BN on a much bigger dataset with large number of subjects. Data preprocessing is done as following

Fig. 5: Face identification results for IJB-A split1 on close protocol. The first column shows the query images from probe templates. The remaining 5 columns show the corresponding top-5 queried gallery templates.

steps. We use OpenCV to detect face and utilize bounding box to crop and resize face images. Each image is 256×256. There are 582,405 images can not be detected, so we delete them. The overlap subject with IJB-A dataset should be removed. Considering the data distribution, we only keep those identities which have 40-500 images. Finally, we obtain 4,356,052 images and 53,317 subjects in total. Our implementation for our face data on GoogLeNet-BN is implemented by caffe [10]. A 224×224 crop is randomly sampled from 256×256 or its horizontal flip. We initialize the weights as in [21] and train GoogLeNet from scratch. SGD with a mini-batch size of 256 is utilized on our GPU cluster machine. The learning rate starts from 0.1 and exp policy is adopted. The weight decay is 0.0001 and the momentum is 0.9. The model are trained for up to 60×10^4 iterations. We stop training procedure when the error is not decreasing.

B. Template-based unconstrained face recognition

After finish training procedure of two DCNN models in source domain. Weights and biases of ResNext 50 and GoogLeNet-BN are shared into target domain. Each face image or frame of video from target domain is viewed as input to feed into those two models, respectively. For ResNext 50, the penultimate global average pooling layer is served as feature extraction layer. It has 2,048 output size. Thus, the feature dimension is 2,048. Given an image or frame $\mathbf{x}_i \in \mathbb{R}^d$ from a mini-batch of size M, where d is the dimension of image or frame. $f_R(\mathbf{x}_i) \in \mathbb{R}^{d_1}$ denotes the feature from ResNext 50, where $d_1 < d$ and $d_1 = 2048$. Similarly, for GoogLeNet-BN, 7×7 average pooling layer is treated as feature extraction layer. The channel size is 1,024. So, the feature dimension is 1,024. Let $f_G(\mathbf{x}_i) \in \mathbb{R}^{d_2}$ is the feature from GoogLenet-BN, where $d_2 = 1024$. In the first-stage fusion, $f_R(\mathbf{x}_i)$ and $f_G(\mathbf{x}_i)$ are concatenated into

(a) The best mated template pairs (b) The worst mated template pairs

Fig. 6: Verification results analysis for mated template pairs on IJB-A split1 .

$f_F(\mathbf{x}_i) \in \mathbb{R}^{d_3}$, where $d_3 = 3072$. Finally, each feature is normalized to unit via L_2 norm for the next procedure.

After feature fusion, in order to train a more discriminative model in target domain, template specific one-vs-rest SVMs play an important role. Specifically, the weights and biases terms for template specific SVMs are learned by optimizing the following L_2-regularized L_2-loss objective function:

$$\begin{aligned}\min_{\mathbf{w}} \frac{1}{2}\mathbf{w}^T\mathbf{w} + \lambda_+ \sum_{i=1}^{N_+} \max\left[0, 1 - y_i \mathbf{w}^T f_F(\mathbf{x}_i)\right]^2 \\ +\lambda_- \sum_{i=1}^{N_-} \max\left[0, 1 - y_j \mathbf{w}^T f_F(\mathbf{x}_j)\right]^2\end{aligned} \tag{1}$$

where $\mathbf{w}$ denote the weights including bias term, $y_i \in \{-1, 1\}$ denotes the label indicating whether the current sample being negative or possible, N_+ indicates the number of positive samples, N_- is the number of negative ones, $N_- \gg N_+$. Moreover, the constraint for negative samples $\lambda_- = C\frac{N_+ + N_-}{2N_-}$, the constraint for positive samples $\lambda_+ = C\frac{N_+ + N_-}{2N_+}$, where C is a trade-off factor. A template includes images or/and frames of video. For the feature of video frame, we compute the average media encodings. Let t_j^V denotes average media encoding of video j.

$$t_j^V = \frac{1}{N_j^V}\sum_{i=1}^{N_j^V} f_F(\mathbf{x}_i) \tag{2}$$

where N_j^V is the number of frame in video j, $\mathbf{x}_i$ denotes i frame of video j. In other words, all features of video frames are aggregate one feature. Thus, the deep facial representations for the ath template can be expressed as

$$T_a = \{t_i^I, ..., t_{N_a}^V\} \tag{3}$$

where t_i^I denotes ith image, N_a express the number of image and video. All media encoding need to perform unit normalization. For verification (a.k.a 1:1 compare), the positive sample of template specific SVM is probe template, the large-scale negative samples include the whole training set. For identification (a.k.a 1:N search), the probe template specific SVMs adopt the whole training set as the large-scale negative samples; whereas for gallery template specific SVM, we adopt other gallery templates and the whole training set as large-scale negative samples. Based on One shot similarity (OSS), we compute similarity between two features p and q via $s(p,q) = \frac{1}{2}\mathcal{P}(q) + \frac{1}{2}\mathcal{Q}(p)$ where $\mathcal{P}(q)$ denotes the trained probe template specific SVM model and $\mathcal{Q}(p)$ indicates the trained gallery template specific SVM model. One template

TABLE I: Performance evaluation on the IJB-A dataset. For 1:1 verification, the true accept rates (TAR) @ false positive rates (FAR) are presented. For 1:N identification, the true positive identification rate (TPIR) @ false positive identification rate (FPIR) and CMC are reported

Method	1:1 Verification TAR			1:N Identification TPIR				
	FAR=0.001	FAR=0.01	FAR=0.1	FPIR=0.01	FPIR=0.1	Rank 1	Rank 5	Rank 10
OpenBR[5]	0.104±0.014	0.236±0.009	0.433±0.006	0.066±0.017	0.149±0.028	0.246±0.011	0.375±0.008	-
GOTS[20]	0.198±0.008	0.406±0.014	0.627±0.012	0.047±0.024	0.235±0.033	0.433±0.021	0.595±0.020	-
B-CNN[30]	-	-	-	0.143±0.027	0.341±0.032	0.588±0.020	0.796±0.017	-
Pooling faces[45]	-	0.309	0.631	-	-	0.846	0.933	0.951
LSFS[23]	0.514±0.060	0.733±0.034	0.895±0.013	0.383±0.063	0.613±0.032	0.820±0.024	0.929±0.013	-
Deep Multi-pose[44]	-	0.787	0.911	0.52	0.75	0.846	0.927	0.947
DCNN*manual*+metric[24]	-	0.787±0.043	0.947±0.011	-	-	0.852±0.018	0.937±0.010	0.954±0.007
Triplet Similarity[31]	0.590±0.050	0.790±0.030	0.945±0.002	0.556±0.065	0.754±0.014	0.880±0.015	0.950±0.007	0.974±0.006
VGG-Face[14]	-	0.805±0.030	-	0.461±0.077	0.670±0.031	0.913±0.011	-	0.981±0.005
PAMs[43]	0.652±0.037	0.826±0.018	-	-	-	0.840±0.012	0.925±0.008	0.946±0.007
DCNN*fusion*[27]	-	0.838±0.042	0.967±0.009	0.577±0.094	0.790±0.033	0.903±0.012	0.965±0.008	0.977±0.007
Masi *et al.*[33]	0.725	0.886	-	-	-	0.906	0.962	0.977
Triplet Embedding[42]	0.813±0.020	0.900±0.010	0.964±0.005	0.753±0.030	0.863±0.014	0.932±0.010	-	0.977±0.005
Template Adaptation[34]	0.836±0.027	0.939±0.013	0.979±0.004	0.774±0.049	0.882±0.016	0.928±0.010	0.977±0.004	0.986±0.003
All-In-One+TPE[46]	0.823±0.020	0.922±0.010	0.976±0.004	0.792±0.020	0.887±0.014	0.947±0.008	-	0.988±0.003
NAN[47]	0.881±0.011	0.941±0.008	0.978±0.003	0.817±0.041	0.917±0.009	0.958±0.005	0.980±0.005	0.986±0.003
L_2-softmax[49]	0.906±0.016	0.952±0.007	0.981±0.003	0.852±0.042	0.930±0.010	0.963±0.007	-	0.986±0.002
L_2-softmax[49]+TPE[42]	0.910±0.013	0.951±0.006	0.979±0.003	0.873±0.024	0.931±0.010	0.961±0.007	-	0.983±0.003
TDFF	0.919±0.006	**0.961±0.007**	0.988±0.003	0.878±0.035	**0.941±0.010**	**0.964±0.006**	**0.988±0.003**	**0.992±0.002**
TDFF+TPE[42]	**0.921±0.005**	**0.961±0.007**	**0.989±0.003**	**0.881±0.039**	**0.940±0.009**	**0.964±0.007**	**0.988±0.003**	**0.992±0.003**

exists many features as Eqn.3, the resulting multiple matching scores should be ensembled into a single one for each template pair in second-stage fusion.

$$s\left(T_a, T_b\right) = \frac{\sum_{t_i \in T_a, t_j \in T_b} s\left(t_i, t_j\right) e^{\beta\, s(t_i,t_j)}}{\sum_{t_i \in T_a, t_j \in T_b} e^{\beta\, s(t_i,t_j)}} \tag{4}$$

where $\beta = 0$ is enough in our following experiments.

TABLE II: Performance evaluation on the IJB-A dataset. For 1:1 verification, the true accept rates (TAR) @ false positive rates (FAR) are presented.

Method	1:1 Verification TAR
	FAR=0.0001
L_2-softmax[49]	0.832±0.027
L_2-softmax[49]+TPE[42]	0.863±0.012
TDFF	**0.875±0.013**
TDFF+TPE[42]	**0.877±0.018**

IV. Experiments and Analysis

In this section, we describe the results for evaluation of the experimental system on the IJB-A verification and identification protocols. The IJB-A dataset contains face images and video frames captured from unconstrained settings which are aligned better with the requirements of real applications. There are 500 subjects with 5,396 images and 2,042 videos sampled to 20,412 frames in total. Full pose variation and wide variations in imaging conditions are the main features of IJB-A dataset, which makes the face recognition very challenging. In our experiments, we just utilize the ground-truth bounding box to crop face image from the original one and resize to 224×224 for each image or frame. We do not use any off-the-shelf pre-trained DCNN model to clean data. We also do not deploy any face detector and do not perform any face alignment procedure.

A remarkable feature of this dataset is that the concept of template is introduced. Each training an testing sample in called a template which comprises a mixture of static images and sampled video frames. Each static image or a frame of video corresponds with a media. On average, each subject has 11.4 images and 4.2 videos. There are 10 training and testing splits. Each of them contains 333 and 167 subjects, respectively.

In TableI, we list the performance of state-of-the-art algorithms on IJB-A dataset. Our performance achieves the best of them for both verification and identification protocols. When we use the TPE to learn a discriminative mapping space while keep the original feature dimension using the training splits of IJB-A. It slightly improves the performance and achieves the new record TAR of 0.921 @ FAR = 0.001, TAR of 0.961 @ FAR = 0.01 and TAR of 0.989 @ FAR = 0.1 for verification. Our method performs significantly better than state-of-the-other algorithms in other indicators as well. These results clearly suggests the effectiveness of our proposed learning framework. In [49], the author reports the results for a very low FAR of 0.0001. Thus, in TableII, we also report the performance @ FAR = 0.0001 for verification protocol, our results still slightly better than L_2-softmax, even TPE is added.

We illustrate the identification results for IJB-A split1 on close protocol in Fig.5. The first column shows the query images from probe templates. The remaining 5 columns show the corresponding top-5 queried gallery templates. For each template, we provide Template ID, Subject ID and similarity score. For all five rows, our approach can successfully find the subjects in rank 1.

Finally, we visualize the verification results in Fig.6 and Fig.7 for IJB-A split1 to gain insight into template based unconstrained face recognition. After computing the similarities for all pairs of probe and reference templates, we sort the resulting list. Each row represents a probe and reference template pair. The original templates within IJB-A contain

(a) The best nonmated template pairs

(b) The worst nonmated template pairs

Fig. 7: Verification results analysis for nonmated template pairs on IJB-A split1 .

from one to dozens of media. Up to eight individual media are shown with the last space showing a mosaic of the remaining media in the template. Between the templates are the Template IDs for probe and reference as well as the best mated and best non-mated similarity. Fig.6 (a) shows the highest mated similarities. In the thirty highest scoring correct matches, we note that every reference template contains dozens of media. The probe templates also contain dozens of media that matches well. Fig.6 (b) shows the lowest mated template pairs, representing failed matching. The thirty lowest mated results from single-media reference templates are under extremely challenging unconstrained conditions. There extremely difficult cases cannot be solved even using our proposed approach. Fig.7 (a) showing the best non-mated similarities shows the most certain nonmates, again often involving large templates with enough guidance from the relevant and historical information. Fig.7 (b) showing the worst non-mated pairs highlights the unstable errors involving single-media reference templates representing impostors in challenging orientation.

V. Conclusion

In this paper, we propose a unified learning framework named transferred deep feature fusion. It can effectively integrate superiority of each module and outperform the state-of-the-art on IJB-A dataset. Inspired by transfer learning, facial feature encoding model of subjects are trained offline in a source domain, and this feature encoding model is transferred to a specific target domain where limited available faces of new subjects can be encoded. Specifically, in order to capture the intrinsic discrimination of subjects and enhance the generalization capability of face recognition models, we deploy two advanced deep convolutional neural networks (DCNN) with distinct architectures to learn the representation of faces on two different large datasets (each one has no overlap with IJB-A dataset) in source domain. These two DCNN models provide distinct feature representations which can better characterize the data distribution from different perspectives. The complementary between two distinct models is beneficial for feature representation. Thus, representing a face from different perspectives could effectively decrease ambiguity among subjects and enhance the generalization performance of face recognition especially on extremely large number of subjects. After offline training procedure, those two DCNN models are transferred to target domain where templates of IJB-A dataset as inputs are performed feature extraction with shared weigths and biases, respectively. Then, two-stage fusion is designed, features from two DCNN models are combined in order to obtain more discriminative representation in first-

stage. Finally, template specific linear SVMs are trained on fused features for classification. Furthermore, for set-to-set matching problem, multiple matching scores are merged into a single one for each template pair as the final results in the second-stage of fusion. Comprehensive evaluations on IJB-A public dataset well demonstrate the significant superiority of the proposed learning framework over other state-of-the-art methods. Based on the proposed approach, we have submitted our IJB-A results to NIST for official evaluation. In the feature, end-to-end network architecture is still attractive for face recognition. Manifold-based metric learning can learn non-linear embedding space, it can explore the geometric structure of the feature encoding.Because, the rotation of head follows a low-dimension manifold. Dictionary learning combines DCNN is an interesting task.

REFERENCES

[1] Pan, Sinno Jialin, and Qiang Yang. "A survey on transfer learning." IEEE Transactions on knowledge and data engineering 22.10 (2010): 1345-1359.
[2] Wolf, Lior, Tal Hassner, and Yaniv Taigman. "Effective unconstrained face recognition by combining multiple descriptors and learned background statistics." IEEE transactions on pattern analysis and machine intelligence 33.10 (2011): 1978-1990.
[3] Chen, Dong, et al. "Bayesian face revisited: A joint formulation." European Conference on Computer Vision. Springer Berlin Heidelberg, 2012.
[4] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems. 2012.
[5] Klontz, Joshua C., et al. "Open source biometric recognition." Biometrics: Theory, Applications and Systems (BTAS), 2013 IEEE Sixth International Conference on. IEEE, 2013.
[6] Lin, Min, Qiang Chen, and Shuicheng Yan. "Network in network." arXiv preprint arXiv:1312.4400 (2013).
[7] Taigman, Yaniv, et al. "Deepface: Closing the gap to human-level performance in face verification." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014.
[8] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014).
[9] Yi, Dong, et al. "Learning face representation from scratch." arXiv preprint arXiv:1411.7923 (2014).
[10] Jia, Yangqing, et al. "Caffe: Convolutional architecture for fast feature embedding." Proceedings of the 22nd ACM international conference on Multimedia. ACM, 2014.
[11] Sharif Razavian, Ali, et al. "CNN features off-the-shelf: an astounding baseline for recognition." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2014.
[12] Sun, Yi, Xiaogang Wang, and Xiaoou Tang. "Deeply learned face representations are sparse, selective, and robust." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015.
[13] Sun, Yi, et al. "Deepid3: Face recognition with very deep neural networks." arXiv preprint arXiv:1502.00873 (2015).
[14] Parkhi, Omkar M., Andrea Vedaldi, and Andrew Zisserman. "Deep Face Recognition." BMVC. Vol. 1. No. 3. 2015.
[15] Srivastava, Rupesh K., Klaus Greff, and Jürgen Schmidhuber. "Training very deep networks." Advances in neural information processing systems. 2015.
[16] Hu, Guosheng, et al. "When face recognition meets with deep learning: an evaluation of convolutional neural networks for face recognition." Proceedings of the IEEE International Conference on Computer Vision Workshops. 2015.
[17] Sainath, Tara N., et al. "Convolutional, long short-term memory, fully connected deep neural networks." Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on. IEEE, 2015.
[18] Szegedy, Christian, et al. "Going deeper with convolutions." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015.
[19] Ioffe, Sergey, and Christian Szegedy. "Batch normalization: Accelerating deep network training by reducing internal covariate shift." arXiv preprint arXiv:1502.03167 (2015).
[20] Klare, Brendan F., et al. "Pushing the frontiers of unconstrained face detection and recognition: IARPA Janus Benchmark A." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015.
[21] He, Kaiming, et al. "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification." Proceedings of the IEEE international conference on computer vision. 2015.
[22] Schroff, Florian, Dmitry Kalenichenko, and James Philbin. "Facenet: A unified embedding for face recognition and clustering." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015.
[23] Wang, Dayong, Charles Otto, and Anil K. Jain. "Face search at scale: 80 million gallery." arXiv preprint arXiv:1507.07242 (2015).
[24] Chen, Jun-Cheng, et al. "An end-to-end system for unconstrained face verification with deep convolutional neural networks." Proceedings of the IEEE International Conference on Computer Vision Workshops. 2015.
[25] Russakovsky, Olga, et al. "Imagenet large scale visual recognition challenge." International Journal of Computer Vision 115.3 (2015): 211-252.
[26] Chen, Tianqi, et al. "Mxnet: A flexible and efficient machine learning library for heterogeneous distributed systems." arXiv preprint arXiv:1512.01274 (2015).
[27] Chen, Jun-Cheng, Vishal M. Patel, and Rama Chellappa. "Unconstrained face verification using deep cnn features." Applications of Computer Vision (WACV), 2016 IEEE Winter Conference on. IEEE, 2016.
[28] Ye, Hao, et al. "Face Recognition via Active Annotation and Learning." Proceedings of the 2016 ACM on Multimedia Conference. ACM, 2016.
[29] Li, Jianshu, et al. "Robust Face Recognition with Deep Multi-View Representation Learning." Proceedings of the 2016 ACM on Multimedia Conference. ACM, 2016.
[30] Chowdhury, Aruni Roy, et al. "One-to-many face recognition with bilinear CNNs." Applications of Computer Vision (WACV), 2016 IEEE Winter Conference on. IEEE, 2016.
[31] Sankaranarayanan, Swami, Azadeh Alavi, and Rama Chellappa. "Triplet similarity embedding for face verification." arXiv preprint arXiv:1602.03418 (2016).
[32] He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016.
[33] Masi, Iacopo, et al. "Do we really need to collect millions of faces for effective face recognition?." European Conference on Computer Vision. Springer International Publishing, 2016.
[34] Crosswhite, Nate, et al. "Template adaptation for face verification and identification." arXiv preprint arXiv:1603.03958 (2016).
[35] Szegedy, Christian, et al. "Rethinking the inception architecture for computer vision." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016.
[36] Szegedy, Christian, et al. "Inception-v4, inception-resnet and the impact of residual connections on learning." arXiv preprint arXiv:1602.07261 (2016).
[37] Wu, Zifeng, Chunhua Shen, and Anton van den Hengel. "Wider or Deeper: Revisiting the ResNet Model for Visual Recognition." arXiv preprint arXiv:1611.10080 (2016).
[38] Targ, Sasha, Diogo Almeida, and Kevin Lyman. "Resnet in Resnet: generalizing residual architectures." arXiv preprint arXiv:1603.08029 (2016).
[39] Xie, Saining, Ross Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He. "Aggregated residual transformations for deep neural networks." arXiv preprint arXiv:1611.05431 (2016).
[40] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Identity mappings in deep residual networks." In European Conference on Computer Vision, pp. 630-645. Springer International Publishing, 2016.
[41] Zagoruyko, Sergey, and Nikos Komodakis. "Wide residual networks." arXiv preprint arXiv:1605.07146 (2016).
[42] Sankaranarayanan, Swami, Azadeh Alavi, Carlos D. Castillo, and Rama Chellappa. "Triplet probabilistic embedding for face verification and clustering." In Biometrics Theory, Applications and Systems (BTAS), 2016 IEEE 8th International Conference on, pp. 1-8. IEEE, 2016.
[43] Masi, Iacopo, Stephen Rawls, Gérard Medioni, and Prem Natarajan. "Pose-aware face recognition in the wild." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4838-4846. 2016.
[44] AbdAlmageed, Wael, Yue Wu, Stephen Rawls, Shai Harel, Tal Hassner, Iacopo Masi, Jongmoo Choi et al. "Face recognition using deep multi-pose representations." In Applications of Computer Vision (WACV), 2016 IEEE Winter Conference on, pp. 1-9. IEEE, 2016.
[45] Hassner, Tal, Iacopo Masi, Jungyeon Kim, Jongmoo Choi, Shai Harel, Prem Natarajan, and Gerard Medioni. "Pooling faces: template based face recognition with pooled face images." In Proceedings of the IEEE

Conference on Computer Vision and Pattern Recognition Workshops, pp. 59-67. 2016.
[46] Ranjan, Rajeev, Swami Sankaranarayanan, Carlos D. Castillo, and Rama Chellappa. "An All-In-One Convolutional Neural Network for Face Analysis." arXiv preprint arXiv:1611.00851 (2016).
[47] Yang, Jiaolong, Peiran Ren, Dong Chen, Fang Wen, Hongdong Li, and Gang Hua. "Neural aggregation network for video face recognition." arXiv preprint arXiv:1603.05474 (2016).
[48] Huang, Gao, Zhuang Liu, Kilian Q. Weinberger, and Laurens van der Maaten. "Densely connected convolutional networks." arXiv preprint arXiv:1608.06993 (2016).
[49] Ranjan, Rajeev and Castillo, Carlos D and Chellappa, Rama. "L2-constrained softmax loss for discriminative face verification." arXiv preprint arXiv:1703.09507v1 (2017).

Lin Xiong received the B.S. degree from Shaanxi University of Science & Technology in 2003, and he received the Ph.D. degree with School of Electronic Engineering, Xidian University, China, in 2014. He is currently an research engineer of Learning & Vision, Core Technology Group, Panasonic R&D Center Singapore, Singapore. His current research interests include face recognition, person re-identification, deep learning, Riemannian manifold optimization, low-rank and sparse matrix factorization, background modeling.

Jayashree Karlekar

Jian Zhao received the B.S. degree from Beihang University in 2012, and he received the Master degree with School of Computer, National University of Defense Technology, China, in 2014. He is currently funded by China Scholarship Council (CSC) and School of Computer, National University of Defense Technology to pursue his Ph.D. degree at Learning and Vision Group, Department of Electronical and Computer Engineering, Faculty of Engineering, National University of Singapore. His current research interests include face recognition, human parsing, human pose estimation, object detection, object semantic segmentation, and relevant deep learning and computer vision problems.

Jiashi Feng is currently an assistant Professor in the department of electrical and computer engineering in the National University of Singapore. He got his B.E. degree from University of Science and Technology, China in 2007 and Ph.D. degree from National University of Singapore in 2014. He was a postdoc researcher at University of California from 2014 to 2015. His current research interest focus on machine learning and computer vision techniques for large-scale data analysis. Specifically, he has done work in object recognition, deep learning, machine learning, highdimensional statistics and big data analysis.

Sugiri Pranata

Shengmei Shen

EXHIBIT R-5

Deep Learning using Linear Support Vector Machines

Yichuan Tang TANG@CS.TORONTO.EDU
Department of Computer Science, University of Toronto. Toronto, Ontario, Canada.

Abstract

Recently, fully-connected and convolutional neural networks have been trained to achieve state-of-the-art performance on a wide variety of tasks such as speech recognition, image classification, natural language processing, and bioinformatics. For classification tasks, most of these "deep learning" models employ the softmax activation function for prediction and minimize cross-entropy loss. In this paper, we demonstrate a small but consistent advantage of replacing the softmax layer with a linear support vector machine. Learning minimizes a margin-based loss instead of the cross-entropy loss. While there have been various combinations of neural nets and SVMs in prior art, our results using L2-SVMs show that by simply replacing softmax with linear SVMs gives significant gains on popular deep learning datasets MNIST, CIFAR-10, and the ICML 2013 Representation Learning Workshop's face expression recognition challenge.

1. Introduction

Deep learning using neural networks have claimed state-of-the-art performances in a wide range of tasks. These include (but not limited to) speech (Mohamed et al., 2009; Dahl et al., 2010) and vision (Jarrett et al., 2009; Ciresan et al., 2011; Rifai et al., 2011a; Krizhevsky et al., 2012). All of the above mentioned papers use the softmax activation function (also known as multinomial logistic regression) for classification.

Support vector machine is an widely used alternative to softmax for classification (Boser et al., 1992). Using SVMs (especially linear) in combination with convolutional nets have been proposed in the past as part of a multistage process. In particular, a deep convolutional net is first trained using supervised/unsupervised objectives to learn good invariant hidden latent representations. The corresponding hidden variables of data samples are then treated as input and fed into linear (or kernel) SVMs (Huang & LeCun, 2006; Lee et al., 2009; Quoc et al., 2010; Coates et al., 2011). This technique usually improves performance but the drawback is that lower level features are not been fine-tuned w.r.t. the SVM's objective.

Other papers have also proposed similar models but with joint training of weights at lower layers using both standard neural nets as well as convolutional neural nets (Zhong & Ghosh, 2000; Collobert & Bengio, 2004; Nagi et al., 2012). In other related works, Weston et al. (2008) proposed a semi-supervised embedding algorithm for deep learning where the hinge loss is combined with the "contrastive loss" from siamese networks (Hadsell et al., 2006). Lower layer weights are learned using stochastic gradient descent. Vinyals et al. (2012) learns a recursive representation using linear SVMs at every layer, but without joint fine-tuning of the hidden representation.

In this paper, we show that for some deep architectures, a linear SVM top layer instead of a softmax is beneficial. We optimize the primal problem of the SVM and the gradients can be backpropagated to learn lower level features. Our models are essentially same as the ones proposed in (Zhong & Ghosh, 2000; Nagi et al., 2012), with the minor novelty of using the loss from the L2-SVM instead of the standard hinge loss. Unlike the hinge loss of a standard SVM, the loss for the L2-SVM is differentiable and penalizes errors much heavily. The primal L2-SVM objective was proposed 3 years before the invention of SVMs (Hinton, 1989)! A similar objective and its optimization are also discussed by (Lee & Mangasarian, 2001).

Compared to nets using a top layer softmax, we demonstrate superior performance on MNIST, CIFAR-10, and on a recent Kaggle competition on recognizing face expressions. Optimization is done using stochastic gradient descent on small minibatches.

International Conference on Machine Learning 2013: Challenges in Representation Learning Workshop. Atlanta, Georgia, USA.

Comparing the two models in Sec. 3.4, we believe the performance gain is largely due to the superior regularization effects of the SVM loss function, rather than an advantage from better parameter optimization.

2. The model

2.1. Softmax

For classification problems using deep learning techniques, it is standard to use the softmax or 1-of-K encoding at the top. For example, given 10 possible classes, the softmax layer has 10 nodes denoted by p_i, where $i = 1, \ldots, 10$. p_i specifies a discrete probability distribution, therefore, $\sum_i^{10} p_i = 1$.

Let $\mathbf{h}$ be the activation of the penultimate layer nodes, $\mathbf{W}$ is the weight connecting the penultimate layer to the softmax layer, the total input into a softmax layer, given by $\mathbf{a}$, is

$$a_i = \sum_k h_k W_{ki}, \tag{1}$$

then we have

$$p_i = \frac{\exp(a_i)}{\sum_j^{10} \exp(a_j)} \tag{2}$$

The predicted class $\hat{i}$ would be

$$\begin{aligned} \hat{i} &= \arg\max_i p_i \\ &= \arg\max_i a_i \end{aligned} \tag{3}$$

2.2. Support Vector Machines

Linear support vector machines (SVM) is originally formulated for binary classification. Given training data and its corresponding labels $(\mathbf{x}_n, y_n)$, $n = 1, \ldots, N$, $\mathbf{x}_n \in \mathbb{R}^D$, $t_n \in \{-1, +1\}$, SVMs learning consists of the following constrained optimization:

$$\begin{aligned} \min_{\mathbf{w}, \xi_n} \quad & \frac{1}{2}\mathbf{w}^\mathsf{T}\mathbf{w} + C\sum_{n=1}^{N} \xi_n \\ s.t. \quad & \mathbf{w}^\mathsf{T}\mathbf{x}_n t_n \geq 1 - \xi_n \quad \forall n \\ & \xi_n \geq 0 \quad \forall n \end{aligned} \tag{4}$$

ξ_n are slack variables which penalizes data points which violate the margin requirements. Note that we can include the bias by augment all data vectors $\mathbf{x}_n$ with a scalar value of 1. The corresponding unconstrained optimization problem is the following:

$$\min_{\mathbf{w}} \quad \frac{1}{2}\mathbf{w}^\mathsf{T}\mathbf{w} + C\sum_{n=1}^{N} \max(1 - \mathbf{w}^\mathsf{T}\mathbf{x}_n t_n, 0) \tag{5}$$

The objective of Eq. 5 is known as the primal form problem of L1-SVM, with the standard hinge loss. Since L1-SVM is not differentiable, a popular variation is known as the L2-SVM which minimizes the squared hinge loss:

$$\min_{\mathbf{w}} \quad \frac{1}{2}\mathbf{w}^\mathsf{T}\mathbf{w} + C\sum_{n=1}^{N} \max(1 - \mathbf{w}^\mathsf{T}\mathbf{x}_n t_n, 0)^2 \tag{6}$$

L2-SVM is differentiable and imposes a bigger (quadratic vs. linear) loss for points which violate the margin. To predict the class label of a test data $\mathbf{x}$:

$$\arg\max_t (\mathbf{w}^\mathsf{T}\mathbf{x})t \tag{7}$$

For Kernel SVMs, optimization must be performed in the dual. However, scalability is a problem with Kernel SVMs, and in this paper we will be only using linear SVMs with standard deep learning models.

2.3. Multiclass SVMs

The simplest way to extend SVMs for multiclass problems is using the so-called *one-vs-rest* approach (Vapnik, 1995). For K class problems, K linear SVMs will be trained independently, where the data from the other classes form the negative cases. Hsu & Lin (2002) discusses other alternative multiclass SVM approaches, but we leave those to future work.

Denoting the output of the k-th SVM as

$$a_k(\mathbf{x}) = \mathbf{w}^\mathsf{T}\mathbf{x} \tag{8}$$

The predicted class is

$$\arg\max_k a_k(\mathbf{x}) \tag{9}$$

Note that prediction using SVMs is exactly the same as using a softmax Eq. 3. The only difference between softmax and multiclass SVMs is in their objectives parametrized by all of the weight matrices $\mathbf{W}$. Softmax layer minimizes cross-entropy or maximizes the log-likelihood, while SVMs simply try to find the maximum margin between data points of different classes.

2.4. Deep Learning with Support Vector Machines

Most deep learning methods for classification using fully connected layers and convolutional layers have used softmax layer objective to learn the lower level parameters. There are exceptions, notably in papers by (Zhong & Ghosh, 2000; Collobert & Bengio, 2004; Nagi et al., 2012), supervised embedding with nonlinear NCA (Salakhutdinov & Hinton, 2007), and semi-supervised deep embedding (Weston et al., 2008). In

this paper, we use L2-SVM's objective to train deep neural nets for classification. Lower layer weights are learned by backpropagating the gradients from the top layer linear SVM. To do this, we need to differentiate the SVM objective with respect to the activation of the penultimate layer. Let the objective in Eq. 5 be $l(\mathbf{w})$, and the input $\mathbf{x}$ is replaced with the penultimate activation $\mathbf{h}$,

$$\frac{\partial l(\mathbf{w})}{\partial \mathbf{h}_n} = -Ct_n\mathbf{w}(\mathbb{I}\{1 > \mathbf{w}^\mathsf{T}\mathbf{h}_n t_n\}) \tag{10}$$

Where $\mathbb{I}\{\cdot\}$ is the indicator function. Likewise, for the L2-SVM, we have

$$\frac{\partial l(\mathbf{w})}{\partial \mathbf{h}_n} = -2Ct_n\mathbf{w}\big(\max(1 - \mathbf{w}^\mathsf{T}\mathbf{h}_n t_n, 0)\big) \tag{11}$$

From this point on, backpropagation algorithm is exactly the same as the standard softmax-based deep learning networks. We found L2-SVM to be slightly better than L1-SVM most of the time and will use the L2-SVM in the experiments section.

3. Experiments

3.1. Facial Expression Recognition

This competition/challenge was hosted by the ICML 2013 workshop on representation learning, organized by the LISA at University of Montreal. The contest itself was hosted on Kaggle with over 120 competing teams during the initial developmental period.

The data consist of 28,709 48x48 images of faces under 7 different types of expression. See Fig 1 for examples and their corresponding expression category. The validation and test sets consist of 3,589 images and this is a classification task.

Winning Solution

We submitted the winning solution with a public validation score of 69.4% and corresponding private test score of 71.2%. Our private test score is almost 2% higher than the 2nd place team. Due to label noise and other factors such as corrupted data, human performance is roughly estimated to be between 65% and 68%[1].

Our submission consists of using a simple Convolutional Neural Network with linear one-vs-all SVM at the top. Stochastic gradient descent with momentum is used for training and several models are averaged to slightly improve the generalization capabilities. Data

Figure 1. Training data. Each column consists of faces of the same expression: starting from the leftmost column: Angry, Disgust, Fear, Happy, Sad, Surprise, Neutral.

preprocessing consisted of first subtracting the mean value of each image and then setting the image norm to be 100. Each pixels is then standardized by removing its mean and dividing its value by the standard deviation of that pixel, across all training images.

Our implementation is in C++ and CUDA, with ports to Matlab using MEX files. Our convolution routines used fast CUDA kernels written by Alex Krizhevsky[2]. The exact model parameters and code is provided on by the author at https://code.google.com/p/deep-learning-faces.

3.1.1. Softmax vs. DLSVM

We compared performances of softmax with the deep learning using L2-SVMs (DLSVM). Both models are tested using an 8 split/fold cross validation, with a image mirroring layer, similarity transformation layer, two convolutional filtering+pooling stages, followed by a fully connected layer with 3072 hidden penultimate hidden units. The hidden layers are all of the rectified linear type. other hyperparameters such as weight decay are selected using cross validation.

We can also look at the validation curve of the Softmax vs L2-SVMs as a function of weight updates in Fig. 2. As learning rate is lowered during the latter half of training, DLSVM maintains a small yet clear performance gain.

We also plotted the 1st layer convolutional filters of the two models:

While not much can be gain from looking at these filters, SVM trained conv net appears to have more

[1]Personal communication from the competition organizers: http://bit.ly/13Zr6Gs

[2]http://code.google.com/p/cuda-convnet

	Softmax	DLSVM L2
Training cross validation	67.6%	68.9%
Public leaderboard	69.3%	69.4%
Private leaderboard	70.1%	71.2%

Table 1. Comparisons of the models in terms of % accuracy. Training c.v. is the average cross validation accuracy over 8 splits. Public leaderboard is the held-out validation set scored via Kaggle's public leaderboard. Private leaderboard is the final private leaderboard score used to determine the competition's winners.

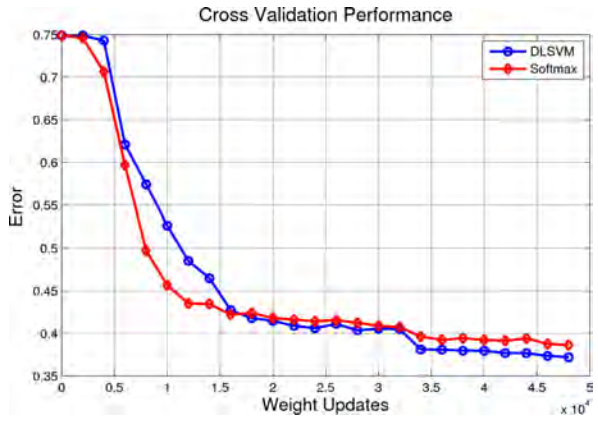


Figure 2. Cross validation performance of the two models. Result is averaged over 8 folds.

textured filters.

3.2. MNIST

MNIST is a standard handwritten digit classification dataset and has been widely used as a benchmark dataset in deep learning. It is a 10 class classification problem with 60,000 training examples and 10,000 test cases.

We used a simple fully connected model by first performing PCA from 784 dimensions down to 70 dimensions. Two hidden layers of 512 units each is followed by a softmax or a L2-SVM. The data is then divided up into 300 minibatches of 200 samples each. We trained using stochastic gradient descent with momentum on these 300 minibatches for over 400 epochs, totaling 120K weight updates. Learning rate is linearly decayed from 0.1 to 0.0. The L2 weight cost on the softmax layer is set to 0.001. To prevent overfitting and critical to achieving good results, a lot of Gaussian noise is added to the input. Noise of standard deviation of 1.0 (linearly decayed to 0) is added. The idea of adding Gaussian noise is taken from these papers (Raiko et al., 2012; Rifai et al., 2011b).

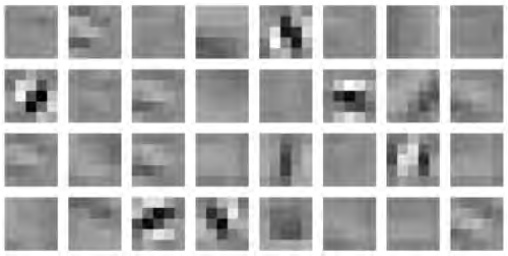


Figure 3. Filters from convolutional net with softmax.

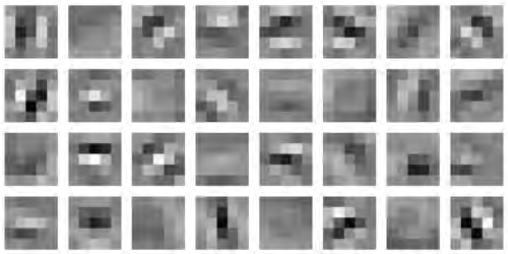


Figure 4. Filters from convolutional net with L2-SVM.

Our learning algorithm is permutation invariant without any unsupervised pretraining and obtains these results: ***Softmax: 0.99%*** ***DLSVM: 0.87%***

An error of 0.87% on MNIST is probably (at this time) state-of-the-art for the above learning setting. The only difference between softmax and DLSVM is the last layer. This experiment is mainly to demonstrate the effectiveness of the last linear SVM layer vs. the softmax, we have not exhaustively explored other commonly used tricks such as Dropout, weight constraints, hidden unit sparsity, adding more hidden layers and increasing the layer size.

3.3. CIFAR-10

Canadian Institute For Advanced Research 10 dataset is a 10 class object dataset with 50,000 images for training and 10,000 for testing. The colored images are 32×32 in resolution. We trained a Convolutional Neural Net with two alternating pooling and filtering layers. Horizontal reflection and jitter is applied to the data randomly before the weight is updated using a minibatch of 128 data cases.

The Convolutional Net part of both the model is fairly standard, the first C layer had 32 5×5 filters with Relu hidden units, the second C layer has 64 5×5 filters. Both pooling layers used max pooling and downsampled by a factor of 2.

The penultimate layer has 3072 hidden nodes and uses

Relu activation with a dropout rate of 0.2. The difference between the Convnet+Softmax and ConvNet with L2-SVM is the mainly in the SVM's C constant, the Softmax's weight decay constant, and the learning rate. We selected the values of these hyperparameters for each model separately using validation.

	ConvNet+Softmax	ConvNet+SVM
Test error	14.0%	11.9%

Table 2. Comparisons of the models in terms of % error on the test set.

In literature, the state-of-the-art (at the time of writing) result is around 9.5% by (Snoeck et al. 2012). However, that model is different as it includes contrast normalization layers as well as used Bayesian optimization to tune its hyperparameters.

3.4. Regularization or Optimization

To see whether the gain in DLSVM is due to the superiority of the objective function or to the ability to better optimize, We looked at the two final models' loss under its own objective functions as well as the other objective. The results are in Table 3.

	ConvNet +Softmax	ConvNet +SVM
Test error	14.0%	11.9%
Avg. cross entropy	0.072	0.353
Hinge loss squared	213.2	0.313

Table 3. Training objective including the weight costs.

It is interesting to note here that lower cross entropy actually led a higher error in the middle row. In addition, we also initialized a ConvNet+Softmax model with the weights of the DLSVM that had 11.9% error. As further training is performed, the network's error rate gradually increased towards 14%.

This gives limited evidence that the gain of DLSVM is largely due to a better objective function.

4. Conclusions

In conclusion, we have shown that DLSVM works better than softmax on 2 standard datasets and a recent dataset. Switching from softmax to SVMs is incredibly simple and appears to be useful for classification tasks. Further research is needed to explore other multiclass SVM formulations and better understand where and how much the gain is obtained.

Acknowledgment

Thanks to Alex Krizhevsky for making his very fast CUDA Conv kernels available! Many thanks to Relu Patrascu for making running experiments possible! Thanks to Ian Goodfellow, Dumitru Erhan, and Yoshua Bengio for organizing the contests.

References

Boser, Bernhard E., Guyon, Isabelle M., and Vapnik, Vladimir N. A training algorithm for optimal margin classifiers. In *Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory*, pp. 144–152. ACM Press, 1992.

Ciresan, D., Meier, U., Masci, J., Gambardella, L. M., and Schmidhuber, J. High-performance neural networks for visual object classification. *CoRR*, abs/1102.0183, 2011.

Coates, Adam, Ng, Andrew Y., and Lee, Honglak. An analysis of single-layer networks in unsupervised feature learning. *Journal of Machine Learning Research - Proceedings Track*, 15:215–223, 2011.

Collobert, R. and Bengio, S. A gentle hessian for efficient gradient descent. In *IEEE International Conference on Acoustic, Speech, and Signal Processing, ICASSP*, 2004.

Dahl, G. E., Ranzato, M., Mohamed, A., and Hinton, G. E. Phone recognition with the mean-covariance restricted Boltzmann machine. In *NIPS 23*. 2010.

Hadsell, Raia, Chopra, Sumit, and Lecun, Yann. Dimensionality reduction by learning an invariant mapping. In *In Proc. Computer Vision and Pattern Recognition Conference (CVPR06*. IEEE Press, 2006.

Hinton, Geoffrey E. Connectionist learning procedures. *Artif. Intell.*, 40(1-3):185–234, 1989.

Hsu, Chih-Wei and Lin, Chih-Jen. A comparison of methods for multiclass support vector machines. *IEEE Transactions on Neural Networks*, 13(2):415–425, 2002.

Huang, F. J. and LeCun, Y. Large-scale learning with SVM and convolutional for generic object categorization. In *CVPR*, pp. I: 284–291, 2006. URL `http://dx.doi.org/10.1109/CVPR.2006.164`.

Jarrett, K., Kavukcuoglu, K., Ranzato, M., and LeCun, Y. What is the best multi-stage architecture for object recognition? In *Proc. Intl. Conf. on Computer Vision (ICCV'09)*. IEEE, 2009.

Krizhevsky, Alex, Sutskever, Ilya, and Hinton, Geoffrey E. Imagenet classification with deep convolutional neural networks. In *NIPS*, pp. 1106–1114, 2012.

Lee, H., Grosse, R., Ranganath, R., and Ng, A. Y. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In *Intl. Conf. on Machine Learning*, pp. 609–616, 2009.

Lee, Yuh-Jye and Mangasarian, O. L. Ssvm: A smooth support vector machine for classification. *Comp. Opt. and Appl.*, 20(1):5–22, 2001.

Mohamed, A., Dahl, G. E., and Hinton, G. E. Deep belief networks for phone recognition. In *NIPS Workshop on Deep Learning for Speech Recognition and Related Applications*, 2009.

Nagi, J., Di Caro, G. A., Giusti, A., , Nagi, F., and Gambardella, L. Convolutional Neural Support Vector Machines: Hybrid visual pattern classifiers for multi-robot systems. In *Proceedings of the 11th International Conference on Machine Learning and Applications (ICMLA)*, Boca Raton, Florida, USA, December 12–15, 2012.

Quoc, L., Ngiam, J., Chen, Z., Chia, D., Koh, P. W., and Ng, A. Tiled convolutional neural networks. In *NIPS 23*. 2010.

Raiko, Tapani, Valpola, Harri, and LeCun, Yann. Deep learning made easier by linear transformations in perceptrons. *Journal of Machine Learning Research - Proceedings Track*, 22:924–932, 2012.

Rifai, Salah, Dauphin, Yann, Vincent, Pascal, Bengio, Yoshua, and Muller, Xavier. The manifold tangent classifier. In *NIPS*, pp. 2294–2302, 2011a.

Rifai, Salah, Glorot, Xavier, Bengio, Yoshua, and Vincent, Pascal. Adding noise to the input of a model trained with a regularized objective. Technical Report 1359, Université de Montréal, Montréal (QC), H3C 3J7, Canada, April 2011b.

Salakhutdinov, Ruslan and Hinton, Geoffrey. Learning a nonlinear embedding by preserving class neighbourhood structure. In *Proceedings of the International Conference on Artificial Intelligence and Statistics*, volume 11, 2007.

Vapnik, V. N. *The nature of statistical learning theory*. Springer, New York, 1995.

Vinyals, O., Jia, Y., Deng, L., and Darrell, T. Learning with Recursive Perceptual Representations. In *NIPS*, 2012.

Weston, Jason, Ratle, Frdric, and Collobert, Ronan. Deep learning via semi-supervised embedding. In *International Conference on Machine Learning*, 2008.

Zhong, Shi and Ghosh, Joydeep. Decision boundary focused neural network classifier. In *Intelligent Engineering Systems Through Articial Neural Networks*, 2000.

EXHIBIT R-6

Notice of Pre-AIA or AIA Status

The present application, filed on or after March 16, 2013, is being examined under the first inventor to file provisions of the AIA.

Response to Amendment

1. Based on applicant's amendment, filed on 3/19/2020, see page 7 through 9, of the remark, with respect to amended claims 1, 4, 6, 10, 13 and newly claims 14-15, have been fully considered and are persuasive, upon further consideration the 35 U.S.C. 112 (b), 35 U.S.C. 112 (a) rejections and rejection of 102(a) (1) for claims 1-15, are hereby withdrawn.

The claims 1-15 are allowed.

REASONS FOR ALLOWANCE

2. The following is an examiner's statement of reasons for allowance.

This invention relates generally, to method for analyzing an image, of a dental arch of a patient, a method in which the analysis image is submitted to a deep learning device, to determine at least one value of a tooth attribute relating to a tooth represented on the analysis image.

Based on applicant's amendment, with respect to claim 1, the closest prior art of record (Kuo), reference is directed to the field of orthodontics. More specifically, the present invention is related to methods and system for providing dynamic orthodontic assessment and treatment profiles, but does not teach or suggest, among other things, "comparison of said image attribute value with a setpoint; sending of an information message as a function of said comparison, the information message being related to the quality of the image acquired, to check whether the analysis image respects the setpoint and, if it does not respect the setpoint, to guide the operator in order for him or her to acquire a new analysis image".

EXHIBIT R-7

Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks

Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun

Abstract—State-of-the-art object detection networks depend on region proposal algorithms to hypothesize object locations. Advances like SPPnet [1] and Fast R-CNN [2] have reduced the running time of these detection networks, exposing region proposal computation as a bottleneck. In this work, we introduce a *Region Proposal Network* (RPN) that shares full-image convolutional features with the detection network, thus enabling nearly cost-free region proposals. An RPN is a fully convolutional network that simultaneously predicts object bounds and objectness scores at each position. The RPN is trained end-to-end to generate high-quality region proposals, which are used by Fast R-CNN for detection. We further merge RPN and Fast R-CNN into a single network by sharing their convolutional features—using the recently popular terminology of neural networks with "attention" mechanisms, the RPN component tells the unified network where to look. For the very deep VGG-16 model [3], our detection system has a frame rate of 5fps (*including all steps*) on a GPU, while achieving state-of-the-art object detection accuracy on PASCAL VOC 2007, 2012, and MS COCO datasets with only 300 proposals per image. In ILSVRC and COCO 2015 competitions, Faster R-CNN and RPN are the foundations of the 1st-place winning entries in several tracks. Code has been made publicly available.

Index Terms—Object Detection, Region Proposal, Convolutional Neural Network.

1 Introduction

Recent advances in object detection are driven by the success of region proposal methods (*e.g.*, [4]) and region-based convolutional neural networks (R-CNNs) [5]. Although region-based CNNs were computationally expensive as originally developed in [5], their cost has been drastically reduced thanks to sharing convolutions across proposals [1], [2]. The latest incarnation, Fast R-CNN [2], achieves near real-time rates using very deep networks [3], *when ignoring the time spent on region proposals*. Now, proposals are the test-time computational bottleneck in state-of-the-art detection systems.

Region proposal methods typically rely on inexpensive features and economical inference schemes. Selective Search [4], one of the most popular methods, greedily merges superpixels based on engineered low-level features. Yet when compared to efficient detection networks [2], Selective Search is an order of magnitude slower, at 2 seconds per image in a CPU implementation. EdgeBoxes [6] currently provides the best tradeoff between proposal quality and speed, at 0.2 seconds per image. Nevertheless, the region proposal step still consumes as much running time as the detection network.

- *S. Ren is with University of Science and Technology of China, Hefei, China. This work was done when S. Ren was an intern at Microsoft Research. Email: sqren@mail.ustc.edu.cn*
- *K. He and J. Sun are with Visual Computing Group, Microsoft Research. E-mail: {kahe,jiansun}@microsoft.com*
- *R. Girshick is with Facebook AI Research. The majority of this work was done when R. Girshick was with Microsoft Research. E-mail: rbg@fb.com*

One may note that fast region-based CNNs take advantage of GPUs, while the region proposal methods used in research are implemented on the CPU, making such runtime comparisons inequitable. An obvious way to accelerate proposal computation is to re-implement it for the GPU. This may be an effective engineering solution, but re-implementation ignores the down-stream detection network and therefore misses important opportunities for sharing computation.

In this paper, we show that an algorithmic change—computing proposals with a deep convolutional neural network—leads to an elegant and effective solution where proposal computation is nearly cost-free given the detection network's computation. To this end, we introduce novel *Region Proposal Networks* (RPNs) that share convolutional layers with state-of-the-art object detection networks [1], [2]. By sharing convolutions at test-time, the marginal cost for computing proposals is small (*e.g.*, 10ms per image).

Our observation is that the convolutional feature maps used by region-based detectors, like Fast R-CNN, can also be used for generating region proposals. On top of these convolutional features, we construct an RPN by adding a few additional convolutional layers that simultaneously regress region bounds and objectness scores at each location on a regular grid. The RPN is thus a kind of fully convolutional network (FCN) [7] and can be trained end-to-end specifically for the task for generating detection proposals.

RPNs are designed to efficiently predict region proposals with a wide range of scales and aspect ratios. In contrast to prevalent methods [8], [9], [1], [2] that use

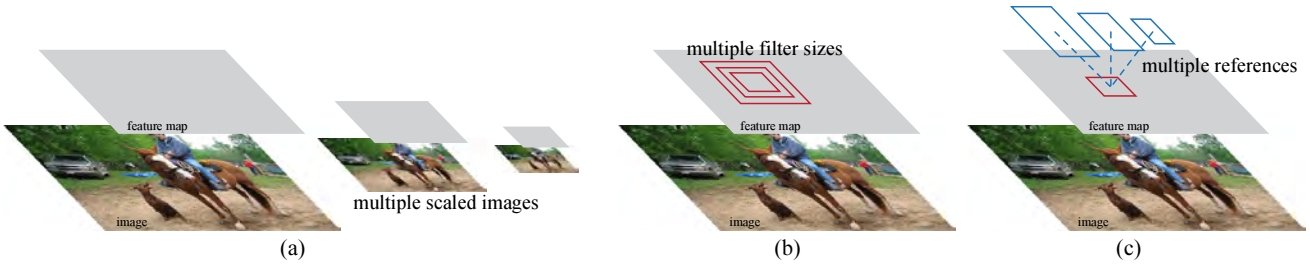


Figure 1: Different schemes for addressing multiple scales and sizes. (a) Pyramids of images and feature maps are built, and the classifier is run at all scales. (b) Pyramids of filters with multiple scales/sizes are run on the feature map. (c) We use pyramids of reference boxes in the regression functions.

pyramids of images (Figure 1, a) or pyramids of filters (Figure 1, b), we introduce novel "anchor" boxes that serve as references at multiple scales and aspect ratios. Our scheme can be thought of as a pyramid of regression references (Figure 1, c), which avoids enumerating images or filters of multiple scales or aspect ratios. This model performs well when trained and tested using single-scale images and thus benefits running speed.

To unify RPNs with Fast R-CNN [2] object detection networks, we propose a training scheme that alternates between fine-tuning for the region proposal task and then fine-tuning for object detection, while keeping the proposals fixed. This scheme converges quickly and produces a unified network with convolutional features that are shared between both tasks.[1]

We comprehensively evaluate our method on the PASCAL VOC detection benchmarks [11] where RPNs with Fast R-CNNs produce detection accuracy better than the strong baseline of Selective Search with Fast R-CNNs. Meanwhile, our method waives nearly all computational burdens of Selective Search at test-time—the effective running time for proposals is just 10 milliseconds. Using the expensive very deep models of [3], our detection method still has a frame rate of 5fps (*including all steps*) on a GPU, and thus is a practical object detection system in terms of both speed and accuracy. We also report results on the MS COCO dataset [12] and investigate the improvements on PASCAL VOC using the COCO data. Code has been made publicly available at `https://github.com/shaoqingren/faster_rcnn` (in MATLAB) and `https://github.com/rbgirshick/py-faster-rcnn` (in Python).

A preliminary version of this manuscript was published previously [10]. Since then, the frameworks of RPN and Faster R-CNN have been adopted and generalized to other methods, such as 3D object detection [13], part-based detection [14], instance segmentation [15], and image captioning [16]. Our fast and effective object detection system has also been built in commercial systems such as at Pinterests [17], with user engagement improvements reported.

In ILSVRC and COCO 2015 competitions, Faster R-CNN and RPN are the basis of several 1st-place entries [18] in the tracks of ImageNet detection, ImageNet localization, COCO detection, and COCO segmentation. RPNs completely learn to propose regions from data, and thus can easily benefit from deeper and more expressive features (such as the 101-layer residual nets adopted in [18]). Faster R-CNN and RPN are also used by several other leading entries in these competitions[2]. These results suggest that our method is not only a cost-efficient solution for practical usage, but also an effective way of improving object detection accuracy.

2 Related Work

Object Proposals. There is a large literature on object proposal methods. Comprehensive surveys and comparisons of object proposal methods can be found in [19], [20], [21]. Widely used object proposal methods include those based on grouping super-pixels (*e.g.*, Selective Search [4], CPMC [22], MCG [23]) and those based on sliding windows (*e.g.*, objectness in windows [24], EdgeBoxes [6]). Object proposal methods were adopted as external modules independent of the detectors (*e.g.*, Selective Search [4] object detectors, R-CNN [5], and Fast R-CNN [2]).

Deep Networks for Object Detection. The R-CNN method [5] trains CNNs end-to-end to classify the proposal regions into object categories or background. R-CNN mainly plays as a classifier, and it does not predict object bounds (except for refining by bounding box regression). Its accuracy depends on the performance of the region proposal module (see comparisons in [20]). Several papers have proposed ways of using deep networks for predicting object bounding boxes [25], [9], [26], [27]. In the OverFeat method [9], a fully-connected layer is trained to predict the box coordinates for the localization task that assumes a single object. The fully-connected layer is then turned

1. Since the publication of the conference version of this paper [10], we have also found that RPNs can be trained jointly with Fast R-CNN networks leading to less training time.

2. http://image-net.org/challenges/LSVRC/2015/results

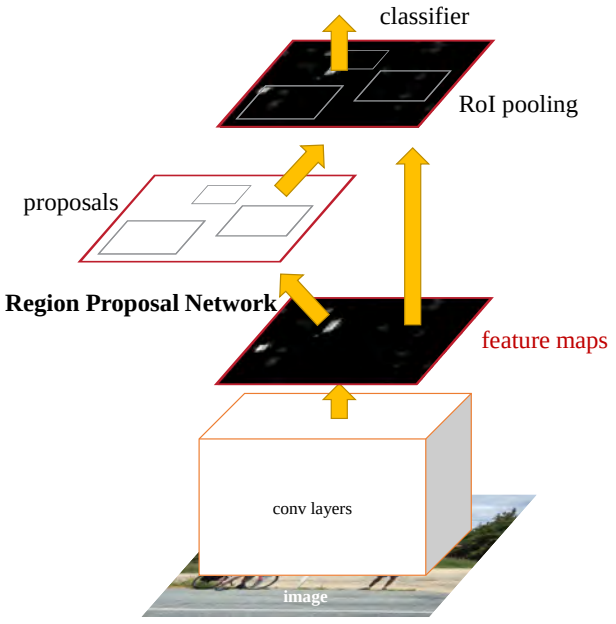


Figure 2: Faster R-CNN is a single, unified network for object detection. The RPN module serves as the 'attention' of this unified network.

into a convolutional layer for detecting multiple class-specific objects. The MultiBox methods [26], [27] generate region proposals from a network whose last fully-connected layer simultaneously predicts multiple class-agnostic boxes, generalizing the "single-box" fashion of OverFeat. These class-agnostic boxes are used as proposals for R-CNN [5]. The MultiBox proposal network is applied on a single image crop or multiple large image crops (*e.g.*, 224×224), in contrast to our fully convolutional scheme. MultiBox does not share features between the proposal and detection networks. We discuss OverFeat and MultiBox in more depth later in context with our method. Concurrent with our work, the DeepMask method [28] is developed for learning segmentation proposals.

Shared computation of convolutions [9], [1], [29], [7], [2] has been attracting increasing attention for efficient, yet accurate, visual recognition. The OverFeat paper [9] computes convolutional features from an image pyramid for classification, localization, and detection. Adaptively-sized pooling (SPP) [1] on shared convolutional feature maps is developed for efficient region-based object detection [1], [30] and semantic segmentation [29]. Fast R-CNN [2] enables end-to-end detector training on shared convolutional features and shows compelling accuracy and speed.

3 Faster R-CNN

Our object detection system, called Faster R-CNN, is composed of two modules. The first module is a deep fully convolutional network that proposes regions, and the second module is the Fast R-CNN detector [2] that uses the proposed regions. The entire system is a single, unified network for object detection (Figure 2). Using the recently popular terminology of neural networks with 'attention' [31] mechanisms, the RPN module tells the Fast R-CNN module where to look. In Section 3.1 we introduce the designs and properties of the network for region proposal. In Section 3.2 we develop algorithms for training both modules with features shared.

3.1 Region Proposal Networks

A Region Proposal Network (RPN) takes an image (of any size) as input and outputs a set of rectangular object proposals, each with an objectness score.[3] We model this process with a fully convolutional network [7], which we describe in this section. Because our ultimate goal is to share computation with a Fast R-CNN object detection network [2], we assume that both nets share a common set of convolutional layers. In our experiments, we investigate the Zeiler and Fergus model [32] (ZF), which has 5 shareable convolutional layers and the Simonyan and Zisserman model [3] (VGG-16), which has 13 shareable convolutional layers.

To generate region proposals, we slide a small network over the convolutional feature map output by the last shared convolutional layer. This small network takes as input an $n \times n$ spatial window of the input convolutional feature map. Each sliding window is mapped to a lower-dimensional feature (256-d for ZF and 512-d for VGG, with ReLU [33] following). This feature is fed into two sibling fully-connected layers—a box-regression layer (*reg*) and a box-classification layer (*cls*). We use $n = 3$ in this paper, noting that the effective receptive field on the input image is large (171 and 228 pixels for ZF and VGG, respectively). This mini-network is illustrated at a single position in Figure 3 (left). Note that because the mini-network operates in a sliding-window fashion, the fully-connected layers are shared across all spatial locations. This architecture is naturally implemented with an $n \times n$ convolutional layer followed by two sibling 1×1 convolutional layers (for *reg* and *cls*, respectively).

3.1.1 Anchors

At each sliding-window location, we simultaneously predict multiple region proposals, where the number of maximum possible proposals for each location is denoted as k. So the *reg* layer has $4k$ outputs encoding the coordinates of k boxes, and the *cls* layer outputs $2k$ scores that estimate probability of object or not object for each proposal[4]. The k proposals are parameterized *relative* to k reference boxes, which we call

3. "Region" is a generic term and in this paper we only consider *rectangular* regions, as is common for many methods (*e.g.*, [27], [4], [6]). "Objectness" measures membership to a set of object classes *vs.* background.

4. For simplicity we implement the *cls* layer as a two-class softmax layer. Alternatively, one may use logistic regression to produce k scores.

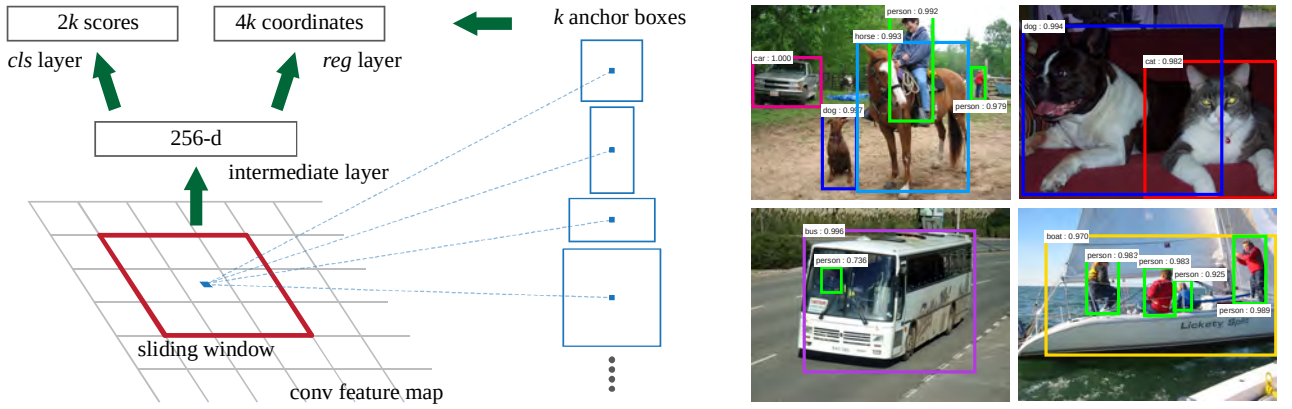


Figure 3: **Left**: Region Proposal Network (RPN). **Right**: Example detections using RPN proposals on PASCAL VOC 2007 test. Our method detects objects in a wide range of scales and aspect ratios.

anchors. An anchor is centered at the sliding window in question, and is associated with a scale and aspect ratio (Figure 3, left). By default we use 3 scales and 3 aspect ratios, yielding $k = 9$ anchors at each sliding position. For a convolutional feature map of a size $W \times H$ (typically ~2,400), there are WHk anchors in total.

Translation-Invariant Anchors

An important property of our approach is that it is *translation invariant*, both in terms of the anchors and the functions that compute proposals relative to the anchors. If one translates an object in an image, the proposal should translate and the same function should be able to predict the proposal in either location. This translation-invariant property is guaranteed by our method[5]. As a comparison, the MultiBox method [27] uses k-means to generate 800 anchors, which are *not* translation invariant. So MultiBox does not guarantee that the same proposal is generated if an object is translated.

The translation-invariant property also reduces the model size. MultiBox has a $(4+1) \times 800$-dimensional fully-connected output layer, whereas our method has a $(4+2) \times 9$-dimensional convolutional output layer in the case of $k = 9$ anchors. As a result, our output layer has 2.8×10^4 parameters ($512 \times (4+2) \times 9$ for VGG-16), two orders of magnitude fewer than MultiBox's output layer that has 6.1×10^6 parameters ($1536 \times (4+1) \times 800$ for GoogleNet [34] in MultiBox [27]). If considering the feature projection layers, our proposal layers still have an order of magnitude fewer parameters than MultiBox[6]. We expect our method to have less risk of overfitting on small datasets, like PASCAL VOC.

5. As is the case of FCNs [7], our network is translation invariant up to the network's total stride.

6. Considering the feature projection layers, our proposal layers' parameter count is $3 \times 3 \times 512 \times 512 + 512 \times 6 \times 9 = 2.4 \times 10^6$; MultiBox's proposal layers' parameter count is $7 \times 7 \times (64+96+64+64) \times 1536 + 1536 \times 5 \times 800 = 27 \times 10^6$.

Multi-Scale Anchors as Regression References

Our design of anchors presents a novel scheme for addressing multiple scales (and aspect ratios). As shown in Figure 1, there have been two popular ways for multi-scale predictions. The first way is based on image/feature pyramids, *e.g.*, in DPM [8] and CNN-based methods [9], [1], [2]. The images are resized at multiple scales, and feature maps (HOG [8] or deep convolutional features [9], [1], [2]) are computed for each scale (Figure 1(a)). This way is often useful but is time-consuming. The second way is to use sliding windows of multiple scales (and/or aspect ratios) on the feature maps. For example, in DPM [8], models of different aspect ratios are trained separately using different filter sizes (such as 5×7 and 7×5). If this way is used to address multiple scales, it can be thought of as a "pyramid of filters" (Figure 1(b)). The second way is usually adopted jointly with the first way [8].

As a comparison, our anchor-based method is built on *a pyramid of anchors*, which is more cost-efficient. Our method classifies and regresses bounding boxes with reference to anchor boxes of multiple scales and aspect ratios. It only relies on images and feature maps of a single scale, and uses filters (sliding windows on the feature map) of a single size. We show by experiments the effects of this scheme for addressing multiple scales and sizes (Table 8).

Because of this multi-scale design based on anchors, we can simply use the convolutional features computed on a single-scale image, as is also done by the Fast R-CNN detector [2]. The design of multi-scale anchors is a key component for sharing features without extra cost for addressing scales.

3.1.2 Loss Function

For training RPNs, we assign a binary class label (of being an object or not) to each anchor. We assign a positive label to two kinds of anchors: (i) the anchor/anchors with the highest Intersection-over-Union (IoU) overlap with a ground-truth box, *or* (ii) an anchor that has an IoU overlap higher than 0.7 with

any ground-truth box. Note that a single ground-truth box may assign positive labels to multiple anchors. Usually the second condition is sufficient to determine the positive samples; but we still adopt the first condition for the reason that in some rare cases the second condition may find no positive sample. We assign a negative label to a non-positive anchor if its IoU ratio is lower than 0.3 for all ground-truth boxes. Anchors that are neither positive nor negative do not contribute to the training objective.

With these definitions, we minimize an objective function following the multi-task loss in Fast R-CNN [2]. Our loss function for an image is defined as:

$$L(\{p_i\},\{t_i\}) = \frac{1}{N_{cls}}\sum_i L_{cls}(p_i, p_i^*) + \lambda\frac{1}{N_{reg}}\sum_i p_i^* L_{reg}(t_i, t_i^*). \quad (1)$$

Here, i is the index of an anchor in a mini-batch and p_i is the predicted probability of anchor i being an object. The ground-truth label p_i^* is 1 if the anchor is positive, and is 0 if the anchor is negative. t_i is a vector representing the 4 parameterized coordinates of the predicted bounding box, and t_i^* is that of the ground-truth box associated with a positive anchor. The classification loss L_{cls} is log loss over two classes (object *vs*. not object). For the regression loss, we use $L_{reg}(t_i, t_i^*) = R(t_i - t_i^*)$ where R is the robust loss function (smooth L_1) defined in [2]. The term $p_i^* L_{reg}$ means the regression loss is activated only for positive anchors ($p_i^* = 1$) and is disabled otherwise ($p_i^* = 0$). The outputs of the *cls* and *reg* layers consist of $\{p_i\}$ and $\{t_i\}$ respectively.

The two terms are normalized by N_{cls} and N_{reg} and weighted by a balancing parameter λ. In our current implementation (as in the released code), the cls term in Eqn.(1) is normalized by the mini-batch size (*i.e.*, $N_{cls} = 256$) and the reg term is normalized by the number of anchor locations (*i.e.*, $N_{reg} \sim 2,400$). By default we set $\lambda = 10$, and thus both *cls* and *reg* terms are roughly equally weighted. We show by experiments that the results are insensitive to the values of λ in a wide range (Table 9). We also note that the normalization as above is not required and could be simplified.

For bounding box regression, we adopt the parameterizations of the 4 coordinates following [5]:

$$\begin{aligned} t_{\mathrm{x}} &= (x - x_{\mathrm{a}})/w_{\mathrm{a}}, \quad t_{\mathrm{y}} = (y - y_{\mathrm{a}})/h_{\mathrm{a}}, \\ t_{\mathrm{w}} &= \log(w/w_{\mathrm{a}}), \quad t_{\mathrm{h}} = \log(h/h_{\mathrm{a}}), \\ t_{\mathrm{x}}^* &= (x^* - x_{\mathrm{a}})/w_{\mathrm{a}}, \quad t_{\mathrm{y}}^* = (y^* - y_{\mathrm{a}})/h_{\mathrm{a}}, \\ t_{\mathrm{w}}^* &= \log(w^*/w_{\mathrm{a}}), \quad t_{\mathrm{h}}^* = \log(h^*/h_{\mathrm{a}}), \end{aligned} \quad (2)$$

where x, y, w, and h denote the box's center coordinates and its width and height. Variables x, x_{a}, and x^* are for the predicted box, anchor box, and ground-truth box respectively (likewise for y, w, h). This can be thought of as bounding-box regression from an anchor box to a nearby ground-truth box.

Nevertheless, our method achieves bounding-box regression by a different manner from previous RoI-based (Region of Interest) methods [1], [2]. In [1], [2], bounding-box regression is performed on features pooled from *arbitrarily* sized RoIs, and the regression weights are *shared* by all region sizes. In our formulation, the features used for regression are of the *same* spatial size (3×3) on the feature maps. To account for varying sizes, a set of k bounding-box regressors are learned. Each regressor is responsible for one scale and one aspect ratio, and the k regressors do *not* share weights. As such, it is still possible to predict boxes of various sizes even though the features are of a fixed size/scale, thanks to the design of anchors.

3.1.3 Training RPNs

The RPN can be trained end-to-end by back-propagation and stochastic gradient descent (SGD) [35]. We follow the "image-centric" sampling strategy from [2] to train this network. Each mini-batch arises from a single image that contains many positive and negative example anchors. It is possible to optimize for the loss functions of all anchors, but this will bias towards negative samples as they are dominate. Instead, we randomly sample 256 anchors in an image to compute the loss function of a mini-batch, where the sampled positive and negative anchors have a ratio of *up to* 1:1. If there are fewer than 128 positive samples in an image, we pad the mini-batch with negative ones.

We randomly initialize all new layers by drawing weights from a zero-mean Gaussian distribution with standard deviation 0.01. All other layers (*i.e.*, the shared convolutional layers) are initialized by pre-training a model for ImageNet classification [36], as is standard practice [5]. We tune all layers of the ZF net, and conv3_1 and up for the VGG net to conserve memory [2]. We use a learning rate of 0.001 for 60k mini-batches, and 0.0001 for the next 20k mini-batches on the PASCAL VOC dataset. We use a momentum of 0.9 and a weight decay of 0.0005 [37]. Our implementation uses Caffe [38].

3.2 Sharing Features for RPN and Fast R-CNN

Thus far we have described how to train a network for region proposal generation, without considering the region-based object detection CNN that will utilize these proposals. For the detection network, we adopt Fast R-CNN [2]. Next we describe algorithms that learn a unified network composed of RPN and Fast R-CNN with shared convolutional layers (Figure 2).

Both RPN and Fast R-CNN, trained independently, will modify their convolutional layers in different ways. We therefore need to develop a technique that allows for sharing convolutional layers between the

Table 1: the learned average proposal size for each anchor using the ZF net (numbers for $s = 600$).

anchor	128^2, 2:1	128^2, 1:1	128^2, 1:2	256^2, 2:1	256^2, 1:1	256^2, 1:2	512^2, 2:1	512^2, 1:1	512^2, 1:2
proposal	188×111	113×114	70×92	416×229	261×284	174×332	768×437	499×501	355×715

two networks, rather than learning two separate networks. We discuss three ways for training networks with features shared:

(i) *Alternating training*. In this solution, we first train RPN, and use the proposals to train Fast R-CNN. The network tuned by Fast R-CNN is then used to initialize RPN, and this process is iterated. This is the solution that is used in all experiments in this paper.

(ii) *Approximate joint training*. In this solution, the RPN and Fast R-CNN networks are merged into one network during training as in Figure 2. In each SGD iteration, the forward pass generates region proposals which are treated just like fixed, pre-computed proposals when training a Fast R-CNN detector. The backward propagation takes place as usual, where for the shared layers the backward propagated signals from both the RPN loss and the Fast R-CNN loss are combined. This solution is easy to implement. But this solution ignores the derivative w.r.t. the proposal boxes' coordinates that are also network responses, so is approximate. In our experiments, we have empirically found this solver produces close results, yet reduces the training time by about 25-50% comparing with alternating training. This solver is included in our released Python code.

(iii) *Non-approximate joint training*. As discussed above, the bounding boxes predicted by RPN are also functions of the input. The RoI pooling layer [2] in Fast R-CNN accepts the convolutional features and also the predicted bounding boxes as input, so a theoretically valid backpropagation solver should also involve gradients w.r.t. the box coordinates. These gradients are ignored in the above approximate joint training. In a non-approximate joint training solution, we need an RoI pooling layer that is differentiable w.r.t. the box coordinates. This is a nontrivial problem and a solution can be given by an "RoI warping" layer as developed in [15], which is beyond the scope of this paper.

4-Step Alternating Training. In this paper, we adopt a pragmatic 4-step training algorithm to learn shared features via alternating optimization. In the first step, we train the RPN as described in Section 3.1.3. This network is initialized with an ImageNet-pre-trained model and fine-tuned end-to-end for the region proposal task. In the second step, we train a separate detection network by Fast R-CNN using the proposals generated by the step-1 RPN. This detection network is also initialized by the ImageNet-pre-trained model. At this point the two networks do not share convolutional layers. In the third step, we use the detector network to initialize RPN training, but we fix the shared convolutional layers and only fine-tune the layers unique to RPN. Now the two networks share convolutional layers. Finally, keeping the shared convolutional layers fixed, we fine-tune the unique layers of Fast R-CNN. As such, both networks share the same convolutional layers and form a unified network. A similar alternating training can be run for more iterations, but we have observed negligible improvements.

3.3 Implementation Details

We train and test both region proposal and object detection networks on images of a single scale [1], [2]. We re-scale the images such that their shorter side is $s = 600$ pixels [2]. Multi-scale feature extraction (using an image pyramid) may improve accuracy but does not exhibit a good speed-accuracy trade-off [2]. On the re-scaled images, the total stride for both ZF and VGG nets on the last convolutional layer is 16 pixels, and thus is ∼10 pixels on a typical PASCAL image before resizing (∼500×375). Even such a large stride provides good results, though accuracy may be further improved with a smaller stride.

For anchors, we use 3 scales with box areas of 128^2, 256^2, and 512^2 pixels, and 3 aspect ratios of 1:1, 1:2, and 2:1. These hyper-parameters are *not* carefully chosen for a particular dataset, and we provide ablation experiments on their effects in the next section. As discussed, our solution does not need an image pyramid or filter pyramid to predict regions of multiple scales, saving considerable running time. Figure 3 (right) shows the capability of our method for a wide range of scales and aspect ratios. Table 1 shows the learned average proposal size for each anchor using the ZF net. We note that our algorithm allows predictions that are larger than the underlying receptive field. Such predictions are not impossible—one may still roughly infer the extent of an object if only the middle of the object is visible.

The anchor boxes that cross image boundaries need to be handled with care. During training, we ignore all cross-boundary anchors so they do not contribute to the loss. For a typical 1000×600 image, there will be roughly 20000 ($\approx 60 \times 40 \times 9$) anchors in total. With the cross-boundary anchors ignored, there are about 6000 anchors per image for training. If the boundary-crossing outliers are not ignored in training, they introduce large, difficult to correct error terms in the objective, and training does not converge. During testing, however, we still apply the fully convolutional RPN to the entire image. This may generate cross-boundary proposal boxes, which we clip to the image boundary.

Table 2: Detection results on **PASCAL VOC 2007 test set** (trained on VOC 2007 trainval). The detectors are Fast R-CNN with ZF, but using various proposal methods for training and testing.

train-time region proposals		**test-time region proposals**		
method	# boxes	method	# proposals	mAP (%)
SS	2000	SS	2000	58.7
EB	2000	EB	2000	58.6
RPN+ZF, shared	2000	RPN+ZF, shared	300	**59.9**
ablation experiments follow below				
RPN+ZF, unshared	2000	RPN+ZF, unshared	300	58.7
SS	2000	RPN+ZF	100	55.1
SS	2000	RPN+ZF	300	56.8
SS	2000	RPN+ZF	1000	56.3
SS	2000	RPN+ZF (no NMS)	6000	55.2
SS	2000	RPN+ZF (no *cls*)	100	44.6
SS	2000	RPN+ZF (no *cls*)	300	51.4
SS	2000	RPN+ZF (no *cls*)	1000	55.8
SS	2000	RPN+ZF (no *reg*)	300	52.1
SS	2000	RPN+ZF (no *reg*)	1000	51.3
SS	2000	RPN+VGG	300	59.2

Some RPN proposals highly overlap with each other. To reduce redundancy, we adopt non-maximum suppression (NMS) on the proposal regions based on their *cls* scores. We fix the IoU threshold for NMS at 0.7, which leaves us about 2000 proposal regions per image. As we will show, NMS does not harm the ultimate detection accuracy, but substantially reduces the number of proposals. After NMS, we use the top-N ranked proposal regions for detection. In the following, we train Fast R-CNN using 2000 RPN proposals, but evaluate different numbers of proposals at test-time.

4 Experiments

4.1 Experiments on PASCAL VOC

We comprehensively evaluate our method on the PASCAL VOC 2007 detection benchmark [11]. This dataset consists of about 5k trainval images and 5k test images over 20 object categories. We also provide results on the PASCAL VOC 2012 benchmark for a few models. For the ImageNet pre-trained network, we use the "fast" version of ZF net [32] that has 5 convolutional layers and 3 fully-connected layers, and the public VGG-16 model[7] [3] that has 13 convolutional layers and 3 fully-connected layers. We primarily evaluate detection mean Average Precision (mAP), because this is the actual metric for object detection (rather than focusing on object proposal proxy metrics).

Table 2 (top) shows Fast R-CNN results when trained and tested using various region proposal methods. These results use the ZF net. For Selective Search (SS) [4], we generate about 2000 proposals by the "fast" mode. For EdgeBoxes (EB) [6], we generate the proposals by the default EB setting tuned for 0.7 IoU. SS has an mAP of 58.7% and EB has an mAP of 58.6% under the Fast R-CNN framework. RPN with Fast R-CNN achieves competitive results, with an mAP of 59.9% while using *up to* 300 proposals[8]. Using RPN yields a much faster detection system than using either SS or EB because of shared convolutional computations; the fewer proposals also reduce the region-wise fully-connected layers' cost (Table 5).

Ablation Experiments on RPN. To investigate the behavior of RPNs as a proposal method, we conducted several ablation studies. First, we show the effect of sharing convolutional layers between the RPN and Fast R-CNN detection network. To do this, we stop after the second step in the 4-step training process. Using separate networks reduces the result slightly to 58.7% (RPN+ZF, unshared, Table 2). We observe that this is because in the third step when the detector-tuned features are used to fine-tune the RPN, the proposal quality is improved.

Next, we disentangle the RPN's influence on training the Fast R-CNN detection network. For this purpose, we train a Fast R-CNN model by using the 2000 SS proposals and ZF net. We fix this detector and evaluate the detection mAP by changing the proposal regions used at test-time. In these ablation experiments, the RPN does not share features with the detector.

Replacing SS with 300 RPN proposals at test-time leads to an mAP of 56.8%. The loss in mAP is because of the inconsistency between the training/testing proposals. This result serves as the baseline for the following comparisons.

Somewhat surprisingly, the RPN still leads to a competitive result (55.1%) when using the top-ranked

7. www.robots.ox.ac.uk/~vgg/research/very_deep/

8. For RPN, the number of proposals (*e.g.*, 300) is the maximum number for an image. RPN may produce fewer proposals after NMS, and thus the average number of proposals is smaller.

Table 3: Detection results on **PASCAL VOC 2007 test set**. The detector is Fast R-CNN and VGG-16. Training data: "07": VOC 2007 trainval, "07+12": union set of VOC 2007 trainval and VOC 2012 trainval. For RPN, the train-time proposals for Fast R-CNN are 2000. †: this number was reported in [2]; using the repository provided by this paper, this result is higher (68.1).

method	# proposals	data	mAP (%)
SS	2000	07	66.9†
SS	2000	07+12	70.0
RPN+VGG, unshared	300	07	68.5
RPN+VGG, shared	300	07	69.9
RPN+VGG, shared	300	07+12	**73.2**
RPN+VGG, shared	300	COCO+07+12	**78.8**

Table 4: Detection results on **PASCAL VOC 2012 test set**. The detector is Fast R-CNN and VGG-16. Training data: "07": VOC 2007 trainval, "07++12": union set of VOC 2007 trainval+test and VOC 2012 trainval. For RPN, the train-time proposals for Fast R-CNN are 2000. †: http://host.robots.ox.ac.uk:8080/anonymous/HZJTQA.html. ‡: http://host.robots.ox.ac.uk:8080/anonymous/YNPLXB.html. §: http://host.robots.ox.ac.uk:8080/anonymous/XEDH10.html.

method	# proposals	data	mAP (%)
SS	2000	12	65.7
SS	2000	07++12	68.4
RPN+VGG, shared†	300	12	67.0
RPN+VGG, shared‡	300	07++12	**70.4**
RPN+VGG, shared§	300	COCO+07++12	**75.9**

Table 5: **Timing** (ms) on a K40 GPU, except SS proposal is evaluated in a CPU. "Region-wise" includes NMS, pooling, fully-connected, and softmax layers. See our released code for the profiling of running time.

model	system	conv	proposal	region-wise	total	rate
VGG	SS + Fast R-CNN	146	1510	174	1830	0.5 fps
VGG	RPN + Fast R-CNN	141	**10**	47	**198**	**5 fps**
ZF	RPN + Fast R-CNN	31	**3**	25	**59**	**17 fps**

100 proposals at test-time, indicating that the top-ranked RPN proposals are accurate. On the other extreme, using the top-ranked 6000 RPN proposals (without NMS) has a comparable mAP (55.2%), suggesting NMS does not harm the detection mAP and may reduce false alarms.

Next, we separately investigate the roles of RPN's *cls* and *reg* outputs by turning off either of them at test-time. When the *cls* layer is removed at test-time (thus no NMS/ranking is used), we randomly sample N proposals from the unscored regions. The mAP is nearly unchanged with $N = 1000$ (55.8%), but degrades considerably to 44.6% when $N = 100$. This shows that the *cls* scores account for the accuracy of the highest ranked proposals.

On the other hand, when the *reg* layer is removed at test-time (so the proposals become anchor boxes), the mAP drops to 52.1%. This suggests that the high-quality proposals are mainly due to the regressed box bounds. The anchor boxes, though having multiple scales and aspect ratios, are not sufficient for accurate detection.

We also evaluate the effects of more powerful networks on the proposal quality of RPN alone. We use VGG-16 to train the RPN, and still use the above detector of SS+ZF. The mAP improves from 56.8% (using RPN+ZF) to 59.2% (using RPN+VGG). This is a promising result, because it suggests that the proposal quality of RPN+VGG is better than that of RPN+ZF. Because proposals of RPN+ZF are competitive with SS (both are 58.7% when consistently used for training and testing), we may expect RPN+VGG to be better than SS. The following experiments justify this hypothesis.

Performance of VGG-16. Table 3 shows the results of VGG-16 for both proposal and detection. Using RPN+VGG, the result is 68.5% for *unshared* features, slightly higher than the SS baseline. As shown above, this is because the proposals generated by RPN+VGG are more accurate than SS. Unlike SS that is pre-defined, the RPN is actively trained and benefits from better networks. For the feature-*shared* variant, the result is 69.9%—better than the strong SS baseline, yet with nearly cost-free proposals. We further train the RPN and detection network on the union set of PASCAL VOC 2007 trainval and 2012 trainval. The mAP is **73.2%**. Figure 5 shows some results on the PASCAL VOC 2007 test set. On the PASCAL VOC 2012 test set (Table 4), our method has an mAP of **70.4%** trained on the union set of VOC 2007 trainval+test and VOC 2012 trainval. Table 6 and Table 7 show the detailed numbers.

Table 6: Results on PASCAL VOC 2007 test set with Fast R-CNN detectors and VGG-16. For RPN, the train-time proposals for Fast R-CNN are 2000. RPN* denotes the unsharing feature version.

method	# box	data	mAP	areo	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	sheep	sofa	train	tv
SS	2000	07	66.9	74.5	78.3	69.2	53.2	36.6	77.3	78.2	82.0	40.7	72.7	67.9	79.6	79.2	73.0	69.0	30.1	65.4	70.2	75.8	65.8
SS	2000	07+12	70.0	77.0	78.1	69.3	59.4	38.3	81.6	78.6	86.7	42.8	78.8	68.9	84.7	82.0	76.6	69.9	31.8	70.1	74.8	80.4	70.4
RPN*	300	07	68.5	74.1	77.2	67.7	53.9	51.0	75.1	79.2	78.9	50.7	78.0	61.1	79.1	81.9	72.2	75.9	37.2	71.4	62.5	77.4	66.4
RPN	300	07	69.9	70.0	80.6	70.1	57.3	49.9	78.2	80.4	82.0	52.2	75.3	67.2	80.3	79.8	75.0	76.3	39.1	68.3	67.3	81.1	67.6
RPN	300	07+12	73.2	76.5	79.0	70.9	65.5	52.1	83.1	84.7	86.4	52.0	81.9	65.7	84.8	84.6	77.5	76.7	38.8	73.6	73.9	83.0	72.6
RPN	300	COCO+07+12	**78.8**	**84.3**	**82.0**	**77.7**	**68.9**	**65.7**	**88.1**	**88.4**	**88.9**	**63.6**	**86.3**	**70.8**	**85.9**	**87.6**	**80.1**	**82.3**	**53.6**	**80.4**	**75.8**	**86.6**	**78.9**

Table 7: Results on PASCAL VOC 2012 test set with Fast R-CNN detectors and VGG-16. For RPN, the train-time proposals for Fast R-CNN are 2000.

method	# box	data	mAP	areo	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	sheep	sofa	train	tv
SS	2000	12	65.7	80.3	74.7	66.9	46.9	37.7	73.9	68.6	87.7	41.7	71.1	51.1	86.0	77.8	79.8	69.8	32.1	65.5	63.8	76.4	61.7
SS	2000	07++12	68.4	82.3	78.4	70.8	52.3	38.7	77.8	71.6	89.3	44.2	73.0	55.0	**87.5**	80.5	80.8	72.0	35.1	68.3	**65.7**	80.4	64.2
RPN	300	12	67.0	82.3	76.4	71.0	48.4	45.2	72.1	72.3	87.3	42.2	73.7	50.0	86.8	78.7	78.4	77.4	34.5	70.1	57.1	77.1	58.9
RPN	300	07++12	70.4	84.9	79.8	74.3	53.9	49.8	77.5	75.9	88.5	45.6	77.1	55.3	86.9	81.7	80.9	79.6	40.1	72.6	60.9	81.2	61.5
RPN	300	COCO+07++12	**75.9**	**87.4**	**83.6**	**76.8**	**62.9**	**59.6**	**81.9**	**82.0**	**91.3**	**54.9**	**82.6**	**59.0**	**89.0**	**85.5**	**84.7**	**84.1**	**52.2**	**78.9**	65.5	**85.4**	**70.2**

Table 8: Detection results of Faster R-CNN on PASCAL VOC 2007 test set using **different settings of anchors**. The network is VGG-16. The training data is VOC 2007 trainval. The default setting of using 3 scales and 3 aspect ratios (69.9%) is the same as that in Table 3.

settings	anchor scales	aspect ratios	mAP (%)
1 scale, 1 ratio	128^2	1:1	65.8
	256^2	1:1	66.7
1 scale, 3 ratios	128^2	{2:1, 1:1, 1:2}	68.8
	256^2	{2:1, 1:1, 1:2}	67.9
3 scales, 1 ratio	$\{128^2, 256^2, 512^2\}$	1:1	**69.8**
3 scales, 3 ratios	$\{128^2, 256^2, 512^2\}$	{2:1, 1:1, 1:2}	**69.9**

Table 9: Detection results of Faster R-CNN on PASCAL VOC 2007 test set using **different values of** λ in Equation (1). The network is VGG-16. The training data is VOC 2007 trainval. The default setting of using $\lambda = 10$ (69.9%) is the same as that in Table 3.

λ	0.1	1	10	100
mAP (%)	67.2	68.9	69.9	69.1

In Table 5 we summarize the running time of the entire object detection system. SS takes 1-2 seconds depending on content (on average about 1.5s), and Fast R-CNN with VGG-16 takes 320ms on 2000 SS proposals (or 223ms if using SVD on fully-connected layers [2]). Our system with VGG-16 takes in total **198ms** for both proposal and detection. With the convolutional features shared, the RPN alone only takes 10ms computing the additional layers. Our region-wise computation is also lower, thanks to fewer proposals (300 per image). Our system has a frame-rate of 17 fps with the ZF net.

Sensitivities to Hyper-parameters. In Table 8 we investigate the settings of anchors. By default we use 3 scales and 3 aspect ratios (69.9% mAP in Table 8). If using just one anchor at each position, the mAP drops by a considerable margin of 3-4%. The mAP is higher if using 3 scales (with 1 aspect ratio) or 3 aspect ratios (with 1 scale), demonstrating that using anchors of multiple sizes as the regression references is an effective solution. Using just 3 scales with 1 aspect ratio (69.8%) is as good as using 3 scales with 3 aspect ratios on this dataset, suggesting that scales and aspect ratios are not disentangled dimensions for the detection accuracy. But we still adopt these two dimensions in our designs to keep our system flexible.

In Table 9 we compare different values of λ in Equation (1). By default we use $\lambda = 10$ which makes the two terms in Equation (1) roughly equally weighted after normalization. Table 9 shows that our result is impacted just marginally (by $\sim 1\%$) when λ is within a scale of about two orders of magnitude (1 to 100). This demonstrates that the result is insensitive to λ in a wide range.

Analysis of Recall-to-IoU. Next we compute the recall of proposals at different IoU ratios with ground-truth boxes. It is noteworthy that the Recall-to-IoU metric is just *loosely* [19], [20], [21] related to the ultimate detection accuracy. It is more appropriate to use this metric to *diagnose* the proposal method than to evaluate it.

In Figure 4, we show the results of using 300, 1000, and 2000 proposals. We compare with SS and EB, and the N proposals are the top-N ranked ones based on the confidence generated by these methods. The plots show that the RPN method behaves gracefully when the number of proposals drops from 2000 to 300. This explains why the RPN has a good ultimate detection mAP when using as few as 300 proposals. As we analyzed before, this property is mainly attributed to the *cls* term of the RPN. The recall of SS and EB drops more quickly than RPN when the proposals are fewer.

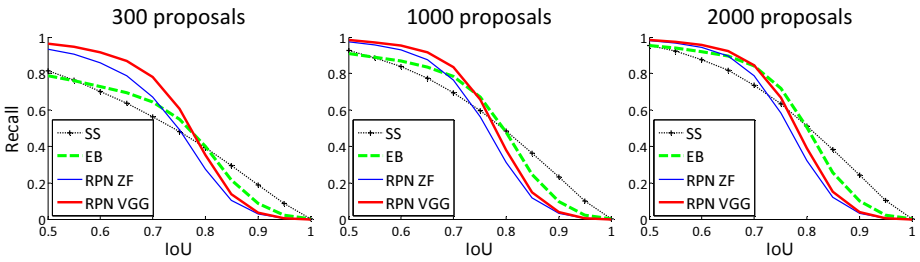


Figure 4: Recall *vs*. IoU overlap ratio on the PASCAL VOC 2007 test set.

Table 10: **One-Stage Detection *vs*. Two-Stage Proposal + Detection**. Detection results are on the PASCAL VOC 2007 test set using the ZF model and Fast R-CNN. RPN uses unshared features.

	proposals		detector	mAP (%)
Two-Stage	RPN + ZF, unshared	300	Fast R-CNN + ZF, 1 scale	58.7
One-Stage	dense, 3 scales, 3 aspect ratios	20000	Fast R-CNN + ZF, 1 scale	53.8
One-Stage	dense, 3 scales, 3 aspect ratios	20000	Fast R-CNN + ZF, 5 scales	53.9

One-Stage Detection *vs*. Two-Stage Proposal + Detection. The OverFeat paper [9] proposes a detection method that uses regressors and classifiers on sliding windows over convolutional feature maps. OverFeat is a *one-stage*, *class-specific* detection pipeline, and ours is a *two-stage cascade* consisting of class-agnostic proposals and class-specific detections. In OverFeat, the region-wise features come from a sliding window of one aspect ratio over a scale pyramid. These features are used to simultaneously determine the location and category of objects. In RPN, the features are from square (3×3) sliding windows and predict proposals relative to anchors with different scales and aspect ratios. Though both methods use sliding windows, the region proposal task is only the first stage of Faster R-CNN—the downstream Fast R-CNN detector *attends* to the proposals to refine them. In the second stage of our cascade, the region-wise features are adaptively pooled [1], [2] from proposal boxes that more faithfully cover the features of the regions. We believe these features lead to more accurate detections.

To compare the one-stage and two-stage systems, we *emulate* the OverFeat system (and thus also circumvent other differences of implementation details) by *one-stage* Fast R-CNN. In this system, the "proposals" are dense sliding windows of 3 scales (128, 256, 512) and 3 aspect ratios (1:1, 1:2, 2:1). Fast R-CNN is trained to predict class-specific scores and regress box locations from these sliding windows. Because the OverFeat system adopts an image pyramid, we also evaluate using convolutional features extracted from 5 scales. We use those 5 scales as in [1], [2].

Table 10 compares the two-stage system and two variants of the one-stage system. Using the ZF model, the one-stage system has an mAP of 53.9%. This is lower than the two-stage system (58.7%) by 4.8%. This experiment justifies the effectiveness of cascaded region proposals and object detection. Similar observations are reported in [2], [39], where replacing SS region proposals with sliding windows leads to ∼6% degradation in both papers. We also note that the one-stage system is slower as it has considerably more proposals to process.

4.2 Experiments on MS COCO

We present more results on the Microsoft COCO object detection dataset [12]. This dataset involves 80 object categories. We experiment with the 80k images on the training set, 40k images on the validation set, and 20k images on the test-dev set. We evaluate the mAP averaged for IoU $\in [0.5 : 0.05 : 0.95]$ (COCO's standard metric, simply denoted as mAP@[.5, .95]) and mAP@0.5 (PASCAL VOC's metric).

There are a few minor changes of our system made for this dataset. We train our models on an 8-GPU implementation, and the effective mini-batch size becomes 8 for RPN (1 per GPU) and 16 for Fast R-CNN (2 per GPU). The RPN step and Fast R-CNN step are both trained for 240k iterations with a learning rate of 0.003 and then for 80k iterations with 0.0003. We modify the learning rates (starting with 0.003 instead of 0.001) because the mini-batch size is changed. For the anchors, we use 3 aspect ratios and 4 scales (adding 64^2), mainly motivated by handling small objects on this dataset. In addition, in our Fast R-CNN step, the negative samples are defined as those with a maximum IoU with ground truth in the interval of $[0, 0.5)$, instead of $[0.1, 0.5)$ used in [1], [2]. We note that in the SPPnet system [1], the negative samples in $[0.1, 0.5)$ are used for network fine-tuning, but the negative samples in $[0, 0.5)$ are still visited in the SVM step with hard-negative mining. But the Fast R-CNN system [2] abandons the SVM step, so the negative samples in $[0, 0.1)$ are never visited. Including these $[0, 0.1)$ samples improves mAP@0.5 on the COCO dataset for both Fast R-CNN and Faster R-CNN systems (but the impact is negligible on PASCAL VOC).

Table 11: Object detection results (%) on the **MS COCO** dataset. The model is VGG-16.

method	proposals	training data	COCO val		COCO test-dev	
			mAP@.5	mAP@[.5, .95]	mAP@.5	mAP@[.5, .95]
Fast R-CNN [2]	SS, 2000	COCO train	-	-	35.9	19.7
Fast R-CNN [impl. in this paper]	SS, 2000	COCO train	38.6	18.9	39.3	19.3
Faster R-CNN	RPN, 300	COCO train	41.5	21.2	42.1	21.5
Faster R-CNN	RPN, 300	COCO trainval	-	-	**42.7**	**21.9**

The rest of the implementation details are the same as on PASCAL VOC. In particular, we keep using 300 proposals and single-scale ($s = 600$) testing. The testing time is still about 200ms per image on the COCO dataset.

In Table 11 we first report the results of the Fast R-CNN system [2] using the implementation in this paper. Our Fast R-CNN baseline has 39.3% mAP@0.5 on the test-dev set, higher than that reported in [2]. We conjecture that the reason for this gap is mainly due to the definition of the negative samples and also the changes of the mini-batch sizes. We also note that the mAP@[.5, .95] is just comparable.

Next we evaluate our Faster R-CNN system. Using the COCO training set to train, Faster R-CNN has 42.1% mAP@0.5 and 21.5% mAP@[.5, .95] on the COCO test-dev set. This is 2.8% higher for mAP@0.5 and **2.2% higher for mAP@[.5, .95]** than the Fast R-CNN counterpart under the same protocol (Table 11). This indicates that RPN performs excellent for improving the localization accuracy at higher IoU thresholds. Using the COCO trainval set to train, Faster R-CNN has 42.7% mAP@0.5 and 21.9% mAP@[.5, .95] on the COCO test-dev set. Figure 6 shows some results on the MS COCO test-dev set.

Faster R-CNN in ILSVRC & COCO 2015 competitions We have demonstrated that Faster R-CNN benefits more from better features, thanks to the fact that the RPN completely learns to propose regions by neural networks. This observation is still valid even when one increases the depth substantially to over 100 layers [18]. Only by replacing VGG-16 with a 101-layer residual net (ResNet-101) [18], the Faster R-CNN system increases the mAP from 41.5%/21.2% (VGG-16) to 48.4%/27.2% (ResNet-101) on the COCO val set. With other improvements orthogonal to Faster R-CNN, He *et al*. [18] obtained a single-model result of 55.7%/34.9% and an ensemble result of 59.0%/37.4% on the COCO test-dev set, which won the 1st place in the COCO 2015 object detection competition. The same system [18] also won the 1st place in the ILSVRC 2015 object detection competition, surpassing the second place by absolute 8.5%. RPN is also a building block of the 1st-place winning entries in ILSVRC 2015 localization and COCO 2015 segmentation competitions, for which the details are available in [18] and [15] respectively.

Table 12: Detection mAP (%) of Faster R-CNN on PASCAL VOC 2007 test set and 2012 test set using different training data. The model is VGG-16. "COCO" denotes that the COCO trainval set is used for training. See also Table 6 and Table 7.

training data	2007 test	2012 test
VOC07	69.9	67.0
VOC07+12	73.2	-
VOC07++12	-	70.4
COCO (no VOC)	76.1	73.0
COCO+VOC07+12	**78.8**	-
COCO+VOC07++12	-	**75.9**

4.3 From MS COCO to PASCAL VOC

Large-scale data is of crucial importance for improving deep neural networks. Next, we investigate how the MS COCO dataset can help with the detection performance on PASCAL VOC.

As a simple baseline, we directly evaluate the COCO detection model on the PASCAL VOC dataset, *without fine-tuning on any PASCAL VOC data*. This evaluation is possible because the categories on COCO are a superset of those on PASCAL VOC. The categories that are exclusive on COCO are ignored in this experiment, and the softmax layer is performed only on the 20 categories plus background. The mAP under this setting is 76.1% on the PASCAL VOC 2007 test set (Table 12). This result is better than that trained on VOC07+12 (73.2%) by a good margin, even though the PASCAL VOC data are not exploited.

Then we fine-tune the COCO detection model on the VOC dataset. In this experiment, the COCO model is in place of the ImageNet-pre-trained model (that is used to initialize the network weights), and the Faster R-CNN system is fine-tuned as described in Section 3.2. Doing so leads to 78.8% mAP on the PASCAL VOC 2007 test set. The extra data from the COCO set increases the mAP by 5.6%. Table 6 shows that the model trained on COCO+VOC has the best AP for every individual category on PASCAL VOC 2007. Similar improvements are observed on the PASCAL VOC 2012 test set (Table 12 and Table 7). We note that the test-time speed of obtaining these strong results is still about **200ms per image**.

5 Conclusion

We have presented RPNs for efficient and accurate region proposal generation. By sharing convolutional

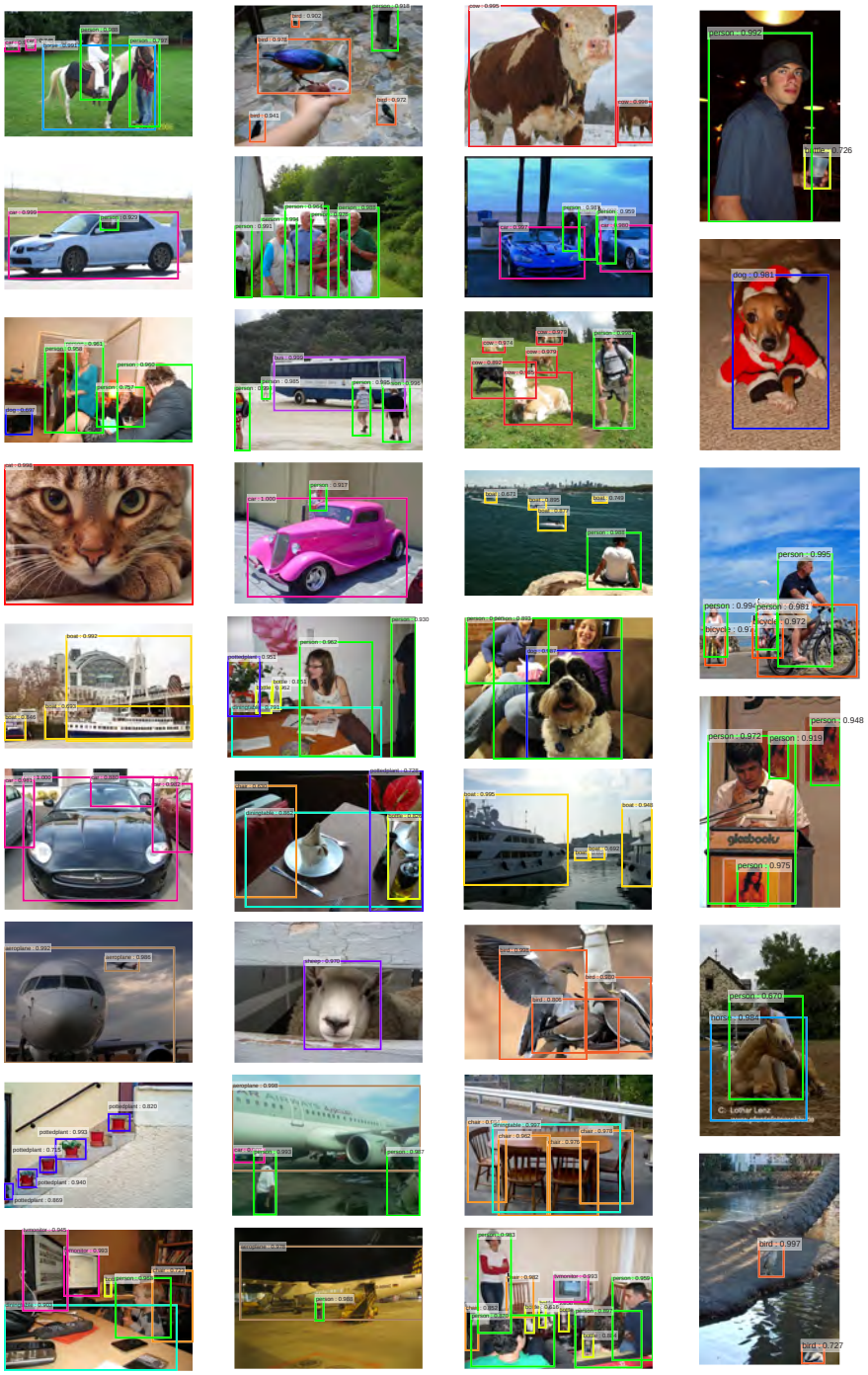


Figure 5: Selected examples of object detection results on the PASCAL VOC 2007 test set using the Faster R-CNN system. The model is VGG-16 and the training data is 07+12 trainval (73.2% mAP on the 2007 test set). Our method detects objects of a wide range of scales and aspect ratios. Each output box is associated with a category label and a softmax score in $[0, 1]$. A score threshold of 0.6 is used to display these images. The running time for obtaining these results is **198ms** per image, *including all steps*.

features with the down-stream detection network, the region proposal step is nearly cost-free. Our method enables a unified, deep-learning-based object detection system to run at near real-time frame rates. The learned RPN also improves region proposal quality and thus the overall object detection accuracy.

REFERENCES

[1] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," in *European Conference on Computer Vision (ECCV)*, 2014.

[2] R. Girshick, "Fast R-CNN," in *IEEE International Conference on Computer Vision (ICCV)*, 2015.

[3] K. Simonyan and A. Zisserman, "Very deep convolutional

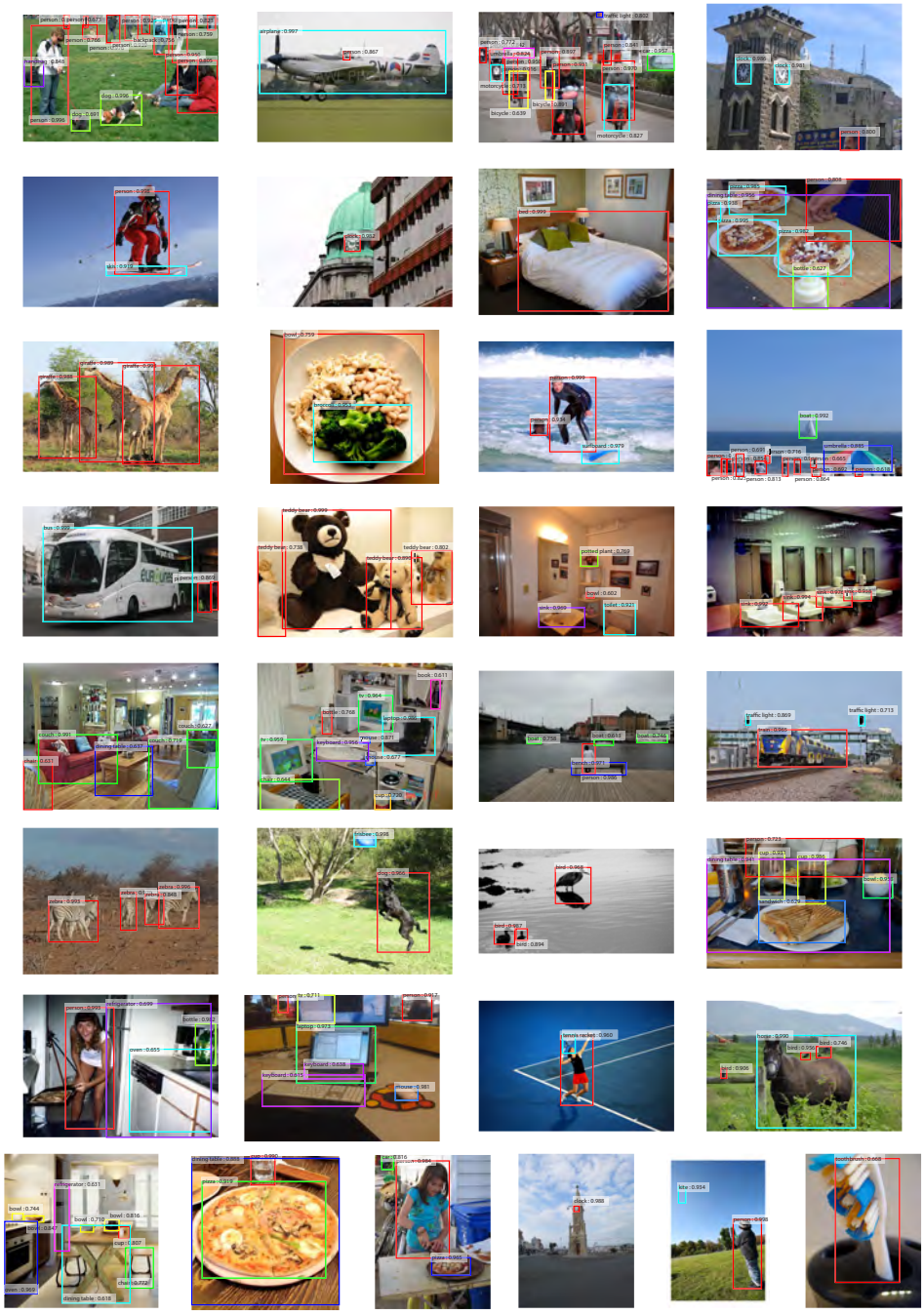


Figure 6: Selected examples of object detection results on the MS COCO test-dev set using the Faster R-CNN system. The model is VGG-16 and the training data is COCO trainval (42.7% mAP@0.5 on the test-dev set). Each output box is associated with a category label and a softmax score in $[0, 1]$. A score threshold of 0.6 is used to display these images. For each image, one color represents one object category in that image.

networks for large-scale image recognition," in *International Conference on Learning Representations (ICLR)*, 2015.

[4] J. R. Uijlings, K. E. van de Sande, T. Gevers, and A. W. Smeulders, "Selective search for object recognition," *International Journal of Computer Vision (IJCV)*, 2013.

[5] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.

[6] C. L. Zitnick and P. Dollár, "Edge boxes: Locating object proposals from edges," in *European Conference on Computer Vision (ECCV)*, 2014.

[7] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

[8] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2010.

[9] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, "Overfeat: Integrated recognition, localization and detection using convolutional networks," in *International Conference on Learning Representations (ICLR)*, 2014.

[10] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards

real-time object detection with region proposal networks," in *Neural Information Processing Systems (NIPS)*, 2015.
[11] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The PASCAL Visual Object Classes Challenge 2007 (VOC2007) Results," 2007.
[12] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: Common Objects in Context," in *European Conference on Computer Vision (ECCV)*, 2014.
[13] S. Song and J. Xiao, "Deep sliding shapes for amodal 3d object detection in rgb-d images," *arXiv:1511.02300*, 2015.
[14] J. Zhu, X. Chen, and A. L. Yuille, "DeePM: A deep part-based model for object detection and semantic part localization," *arXiv:1511.07131*, 2015.
[15] J. Dai, K. He, and J. Sun, "Instance-aware semantic segmentation via multi-task network cascades," *arXiv:1512.04412*, 2015.
[16] J. Johnson, A. Karpathy, and L. Fei-Fei, "Densecap: Fully convolutional localization networks for dense captioning," *arXiv:1511.07571*, 2015.
[17] D. Kislyuk, Y. Liu, D. Liu, E. Tzeng, and Y. Jing, "Human curation and convnets: Powering item-to-item recommendations on pinterest," *arXiv:1511.04003*, 2015.
[18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *arXiv:1512.03385*, 2015.
[19] J. Hosang, R. Benenson, and B. Schiele, "How good are detection proposals, really?" in *British Machine Vision Conference (BMVC)*, 2014.
[20] J. Hosang, R. Benenson, P. Dollár, and B. Schiele, "What makes for effective detection proposals?" *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2015.
[21] N. Chavali, H. Agrawal, A. Mahendru, and D. Batra, "Object-Proposal Evaluation Protocol is 'Gameable'," *arXiv: 1505.05836*, 2015.
[22] J. Carreira and C. Sminchisescu, "CPMC: Automatic object segmentation using constrained parametric min-cuts," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2012.
[23] P. Arbeláez, J. Pont-Tuset, J. T. Barron, F. Marques, and J. Malik, "Multiscale combinatorial grouping," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.
[24] B. Alexe, T. Deselaers, and V. Ferrari, "Measuring the objectness of image windows," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2012.
[25] C. Szegedy, A. Toshev, and D. Erhan, "Deep neural networks for object detection," in *Neural Information Processing Systems (NIPS)*, 2013.
[26] D. Erhan, C. Szegedy, A. Toshev, and D. Anguelov, "Scalable object detection using deep neural networks," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.
[27] C. Szegedy, S. Reed, D. Erhan, and D. Anguelov, "Scalable, high-quality object detection," *arXiv:1412.1441 (v1)*, 2015.
[28] P. O. Pinheiro, R. Collobert, and P. Dollar, "Learning to segment object candidates," in *Neural Information Processing Systems (NIPS)*, 2015.
[29] J. Dai, K. He, and J. Sun, "Convolutional feature masking for joint object and stuff segmentation," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.
[30] S. Ren, K. He, R. Girshick, X. Zhang, and J. Sun, "Object detection networks on convolutional feature maps," *arXiv:1504.06066*, 2015.
[31] J. K. Chorowski, D. Bahdanau, D. Serdyuk, K. Cho, and Y. Bengio, "Attention-based models for speech recognition," in *Neural Information Processing Systems (NIPS)*, 2015.
[32] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional neural networks," in *European Conference on Computer Vision (ECCV)*, 2014.
[33] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *International Conference on Machine Learning (ICML)*, 2010.
[34] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, and A. Rabinovich, "Going deeper with convolutions," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.
[35] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural computation*, 1989.
[36] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," in *International Journal of Computer Vision (IJCV)*, 2015.
[37] A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks," in *Neural Information Processing Systems (NIPS)*, 2012.
[38] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," *arXiv:1408.5093*, 2014.
[39] K. Lenc and A. Vedaldi, "R-CNN minus R," in *British Machine Vision Conference (BMVC)*, 2015.

EXHIBIT R-8

US 20130209954A1

(19) **United States**

(12) **Patent Application Publication**

Prakash et al.

(10) **Pub. No.: US 2013/0209954 A1**

(43) **Pub. Date: Aug. 15, 2013**

(54) **TECHNIQUES FOR STANDARDIZED IMAGING OF ORAL CAVITY**

(71) Applicants: **Manu Prakash**, San Francisco, CA (US); **Dhruv Boddupalli**, Redwood City, CA (US); **James Clements**, E. Palo Alto, CA (US); **Aditya Gande**, Cupertino, CA (US)

(72) Inventors: **Manu Prakash**, San Francisco, CA (US); **Dhruv Boddupalli**, Redwood City, CA (US); **James Clements**, E. Palo Alto, CA (US); **Aditya Gande**, Cupertino, CA (US)

(21) Appl. No.: **13/764,764**

(22) Filed: **Feb. 11, 2013**

Related U.S. Application Data

(60) Provisional application No. 61/597,772, filed on Feb. 11, 2012.

Publication Classification

(51) **Int. Cl.**

A61B 1/24	(2006.01)
A61B 1/06	(2006.01)
A61B 1/00	(2006.01)
A61C 5/14	(2006.01)
A61B 1/04	(2006.01)

(52) **U.S. Cl.**

CPC ... ***A61B 1/24*** (2013.01); ***A61C 5/14*** (2013.01); ***A61B 1/04*** (2013.01); ***A61B 1/00165*** (2013.01); ***A61B 1/00011*** (2013.01); ***A61B 1/00045*** (2013.01); ***A61B 1/00009*** (2013.01); ***A61B 1/0646*** (2013.01); ***A61B 1/0002*** (2013.01); ***A61B 1/00188*** (2013.01)

USPC **433/29**; 433/215

(57) **ABSTRACT**

A method, system and apparatus for imaging an oral cavity of a subject include a bracket comprising a mouthpiece and a camera mount. The mouthpiece has upper and lower bite guides disposed on a posterior side and separated by an opening through the mouthpiece. The bite guides are spaced apart so a subject biting on the guides opens the oral cavity to inspection through the opening. The mount is disposed on an anterior side of the mouthpiece; and has a flange to engage and slide along the opening and an optical path for light to pass through the mount and mouthpiece. A clip on an anterior side of the mount is configured to removeably hold a camera to record light passing through the optical path from the posterior side of the mount. In one embodiment, the camera is a programmable cell phone with digital camera.

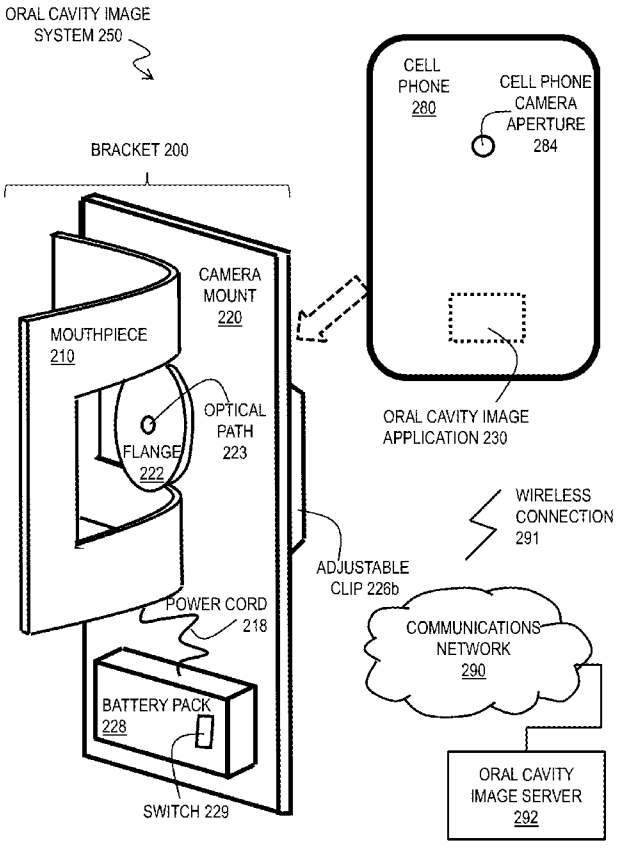
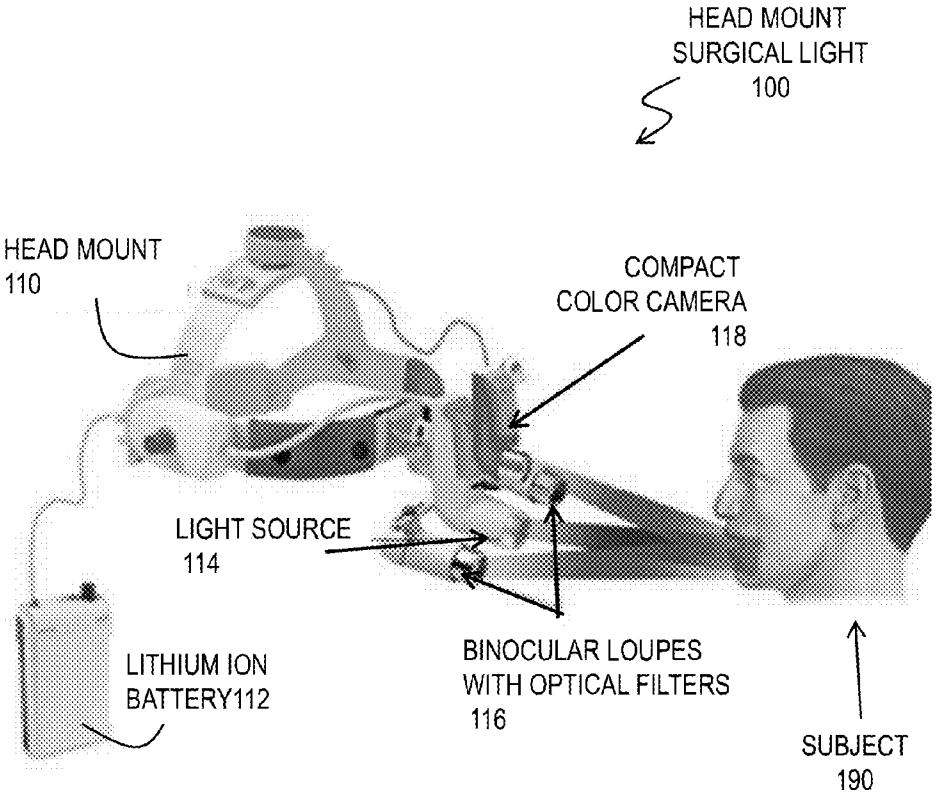


FIG. 1



(PRIOR ART)

FIG. 2A

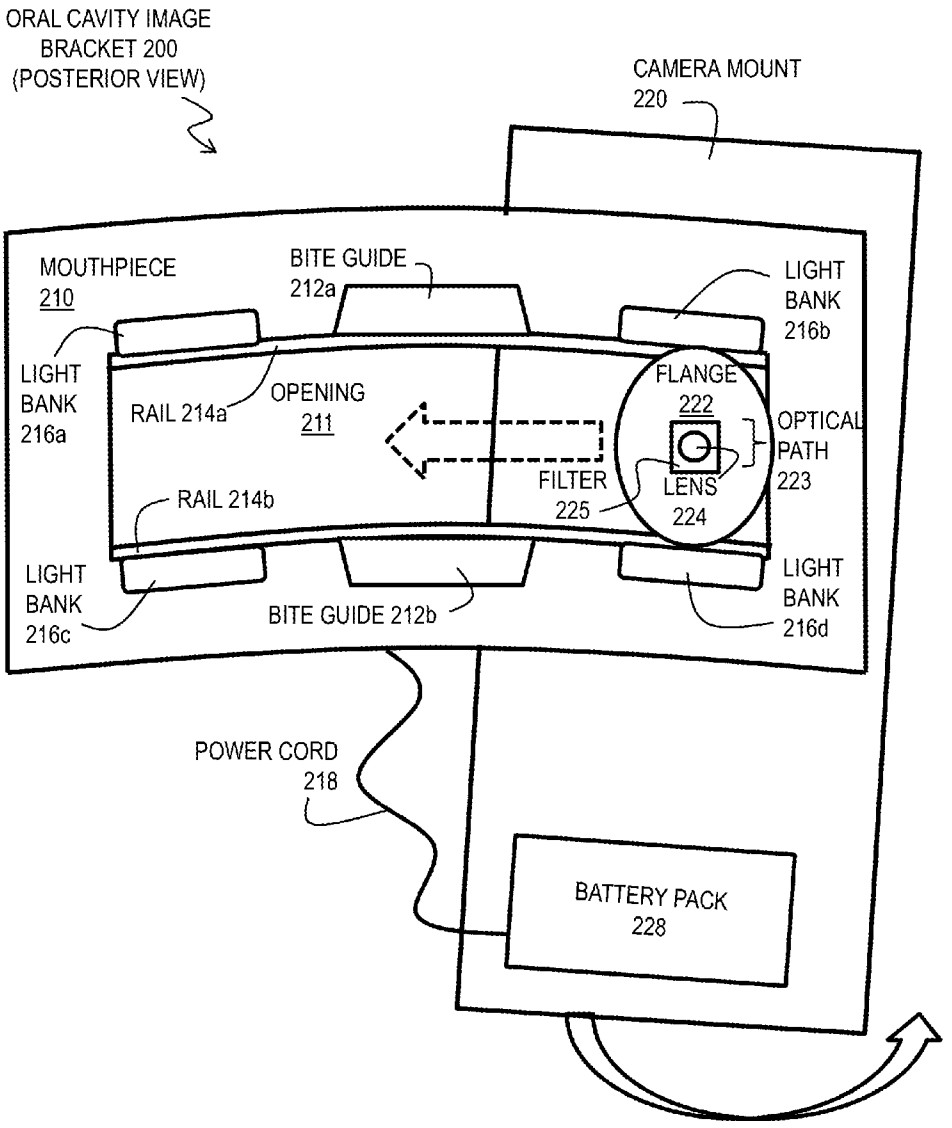

ORAL CAVITY IMAGE BRACKET 200 (POSTERIOR VIEW)
CAMERA MOUNT 220
MOUTHPIECE 210
BITE GUIDE 212a
LIGHT BANK 216b
LIGHT BANK 216a
RAIL 214a
OPENING 211
FLANGE 222
OPTICAL PATH 223
FILTER 225
LENS 224
RAIL 214b
LIGHT BANK 216c
BITE GUIDE 212b
LIGHT BANK 216d
POWER CORD 218
BATTERY PACK 228

FIG. 2B

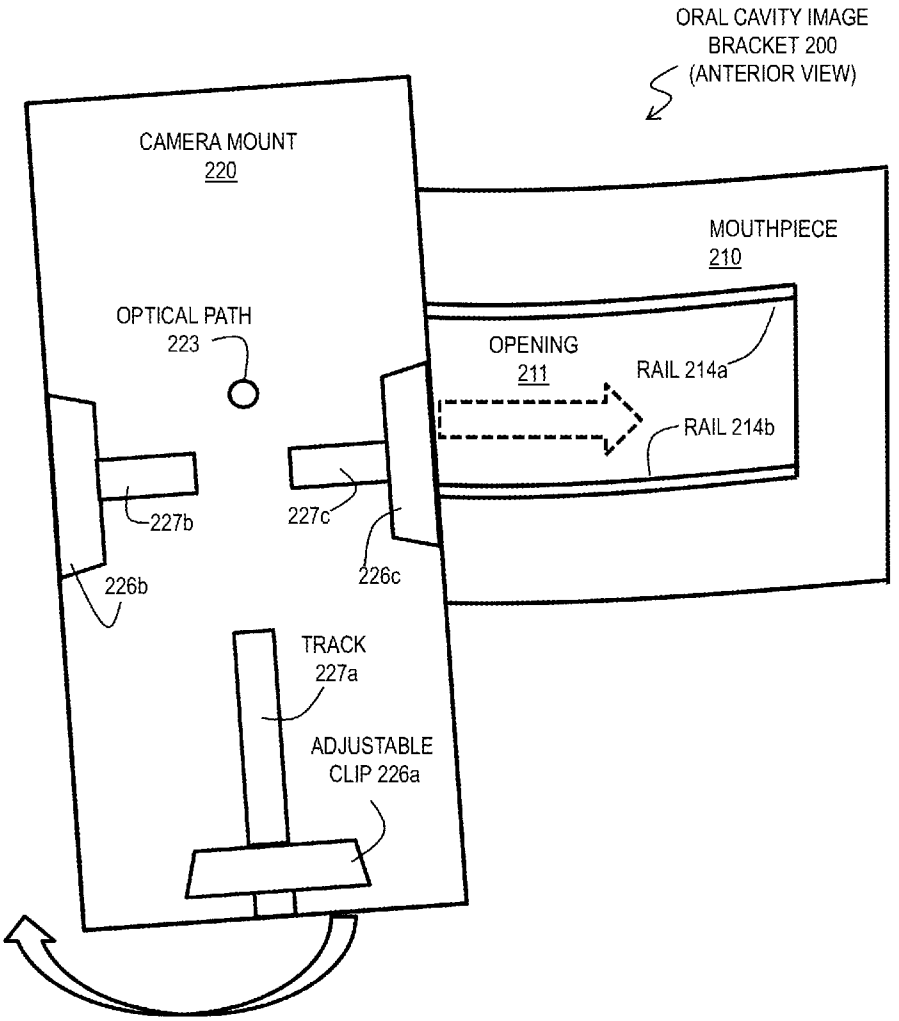

ORAL CAVITY IMAGE
BRACKET 200
(ANTERIOR VIEW)
CAMERA MOUNT
220
MOUTHPIECE
210
OPTICAL PATH
223
OPENING
211
RAIL 214a
RAIL 214b
227b
227c
226b
226c
TRACK
227a
ADJUSTABLE
CLIP 226a

FIG. 2C

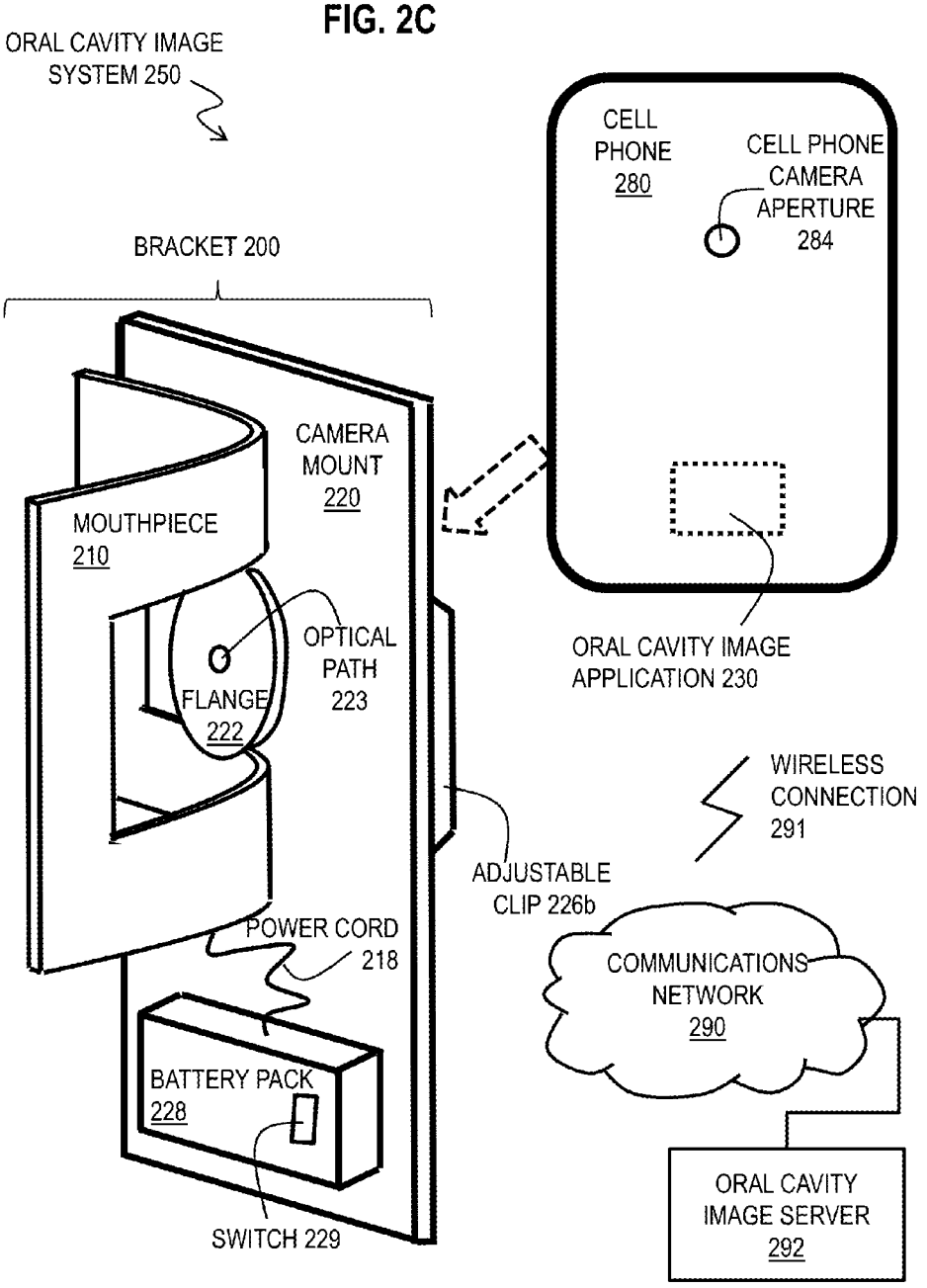

ORAL CAVITY IMAGE SYSTEM 250
CELL PHONE 280
CELL PHONE CAMERA APERTURE 284
BRACKET 200
CAMERA MOUNT 220
MOUTHPIECE 210
OPTICAL PATH 223
FLANGE 222
ORAL CAVITY IMAGE APPLICATION 230
WIRELESS CONNECTION 291
ADJUSTABLE CLIP 226b
POWER CORD 218
COMMUNICATIONS NETWORK 290
BATTERY PACK 228
SWITCH 229
ORAL CAVITY IMAGE SERVER 292

FIG. 3A

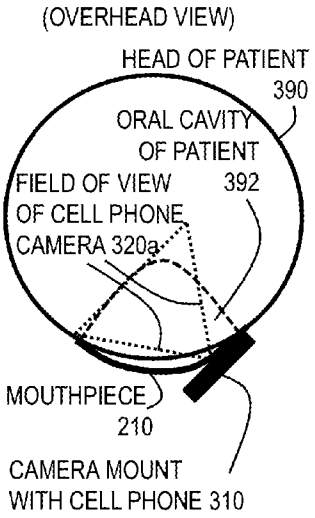


FIG. 3B

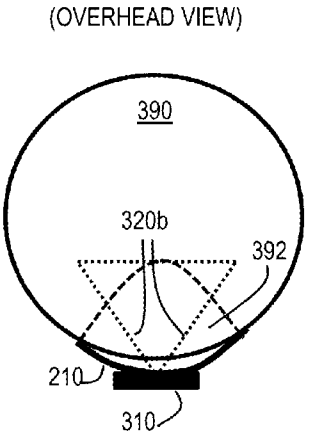


FIG. 3C

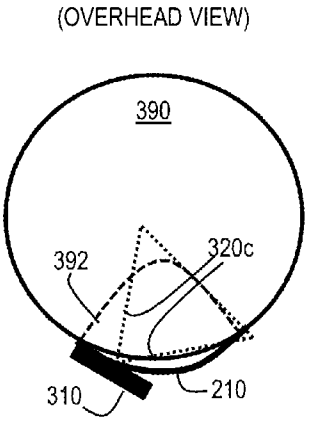


FIG. 3D

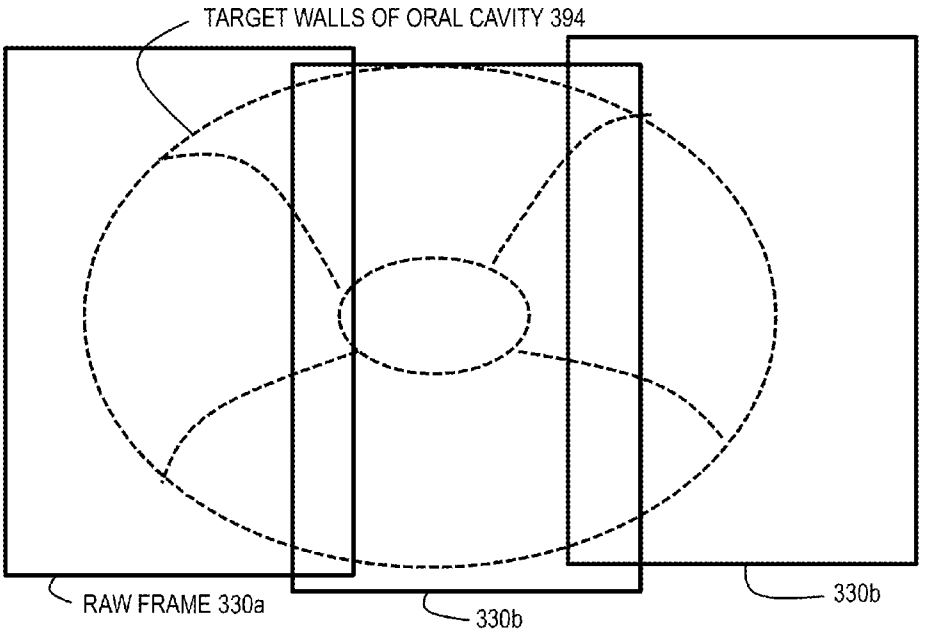


FIG. 4A

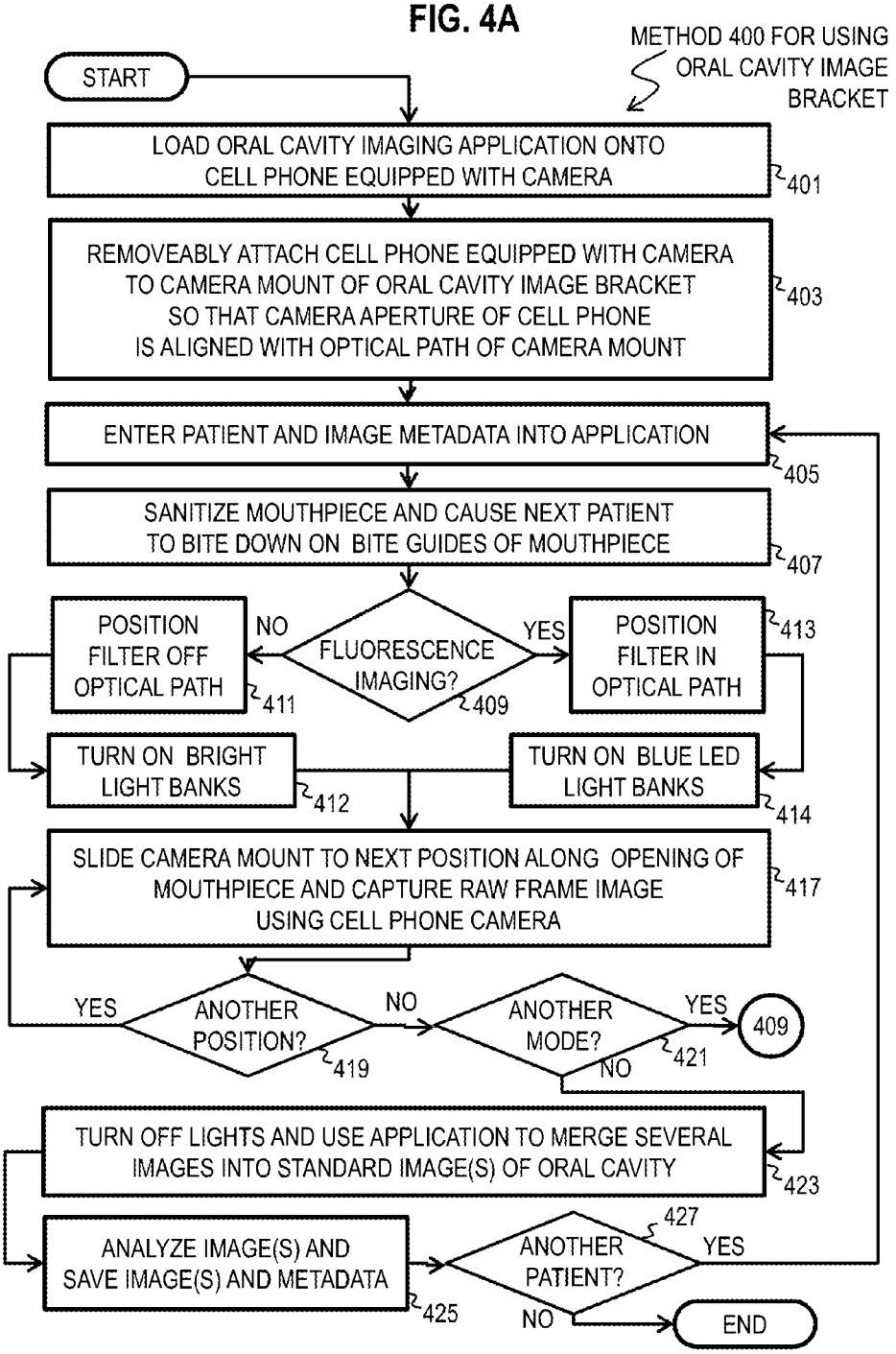

METHOD 400 FOR USING ORAL CAVITY IMAGE BRACKET
START
LOAD ORAL CAVITY IMAGING APPLICATION ONTO CELL PHONE EQUIPPED WITH CAMERA
401
REMOVEABLY ATTACH CELL PHONE EQUIPPED WITH CAMERA TO CAMERA MOUNT OF ORAL CAVITY IMAGE BRACKET SO THAT CAMERA APERTURE OF CELL PHONE IS ALIGNED WITH OPTICAL PATH OF CAMERA MOUNT
403
ENTER PATIENT AND IMAGE METADATA INTO APPLICATION
405
SANITIZE MOUTHPIECE AND CAUSE NEXT PATIENT TO BITE DOWN ON BITE GUIDES OF MOUTHPIECE
407
FLUORESCENCE IMAGING?
409
NO
POSITION FILTER OFF OPTICAL PATH
411
YES
POSITION FILTER IN OPTICAL PATH
413
TURN ON BRIGHT LIGHT BANKS
412
TURN ON BLUE LED LIGHT BANKS
414
SLIDE CAMERA MOUNT TO NEXT POSITION ALONG OPENING OF MOUTHPIECE AND CAPTURE RAW FRAME IMAGE USING CELL PHONE CAMERA
417
YES
ANOTHER POSITION?
419
NO
ANOTHER MODE?
421
YES
409
NO
TURN OFF LIGHTS AND USE APPLICATION TO MERGE SEVERAL IMAGES INTO STANDARD IMAGE(S) OF ORAL CAVITY
423
ANALYZE IMAGE(S) AND SAVE IMAGE(S) AND METADATA
425
ANOTHER PATIENT?
427
YES
NO
END

FIG. 4B

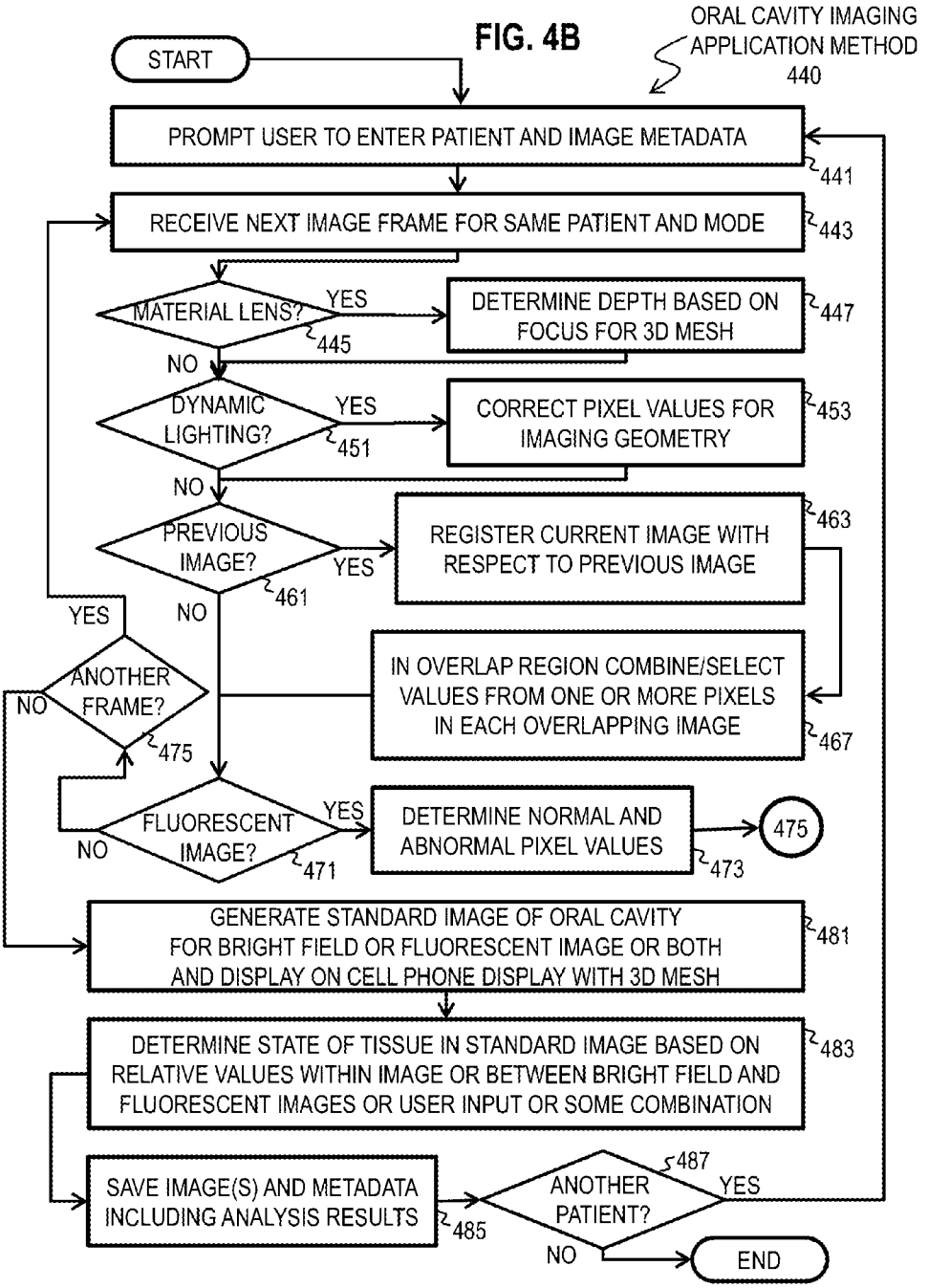

ORAL CAVITY IMAGING APPLICATION METHOD 440
START
PROMPT USER TO ENTER PATIENT AND IMAGE METADATA
441
RECEIVE NEXT IMAGE FRAME FOR SAME PATIENT AND MODE
443
MATERIAL LENS?
445
YES
DETERMINE DEPTH BASED ON FOCUS FOR 3D MESH
447
NO
DYNAMIC LIGHTING?
451
YES
CORRECT PIXEL VALUES FOR IMAGING GEOMETRY
453
NO
PREVIOUS IMAGE?
461
YES
REGISTER CURRENT IMAGE WITH RESPECT TO PREVIOUS IMAGE
463
NO
YES
ANOTHER FRAME?
475
NO
IN OVERLAP REGION COMBINE/SELECT VALUES FROM ONE OR MORE PIXELS IN EACH OVERLAPPING IMAGE
467
FLUORESCENT IMAGE?
471
YES
NO
DETERMINE NORMAL AND ABNORMAL PIXEL VALUES
473
475
GENERATE STANDARD IMAGE OF ORAL CAVITY FOR BRIGHT FIELD OR FLUORESCENT IMAGE OR BOTH AND DISPLAY ON CELL PHONE DISPLAY WITH 3D MESH
481
DETERMINE STATE OF TISSUE IN STANDARD IMAGE BASED ON RELATIVE VALUES WITHIN IMAGE OR BETWEEN BRIGHT FIELD AND FLUORESCENT IMAGES OR USER INPUT OR SOME COMBINATION
483
SAVE IMAGE(S) AND METADATA INCLUDING ANALYSIS RESULTS
485
ANOTHER PATIENT?
487
YES
NO
END

FIG. 5A

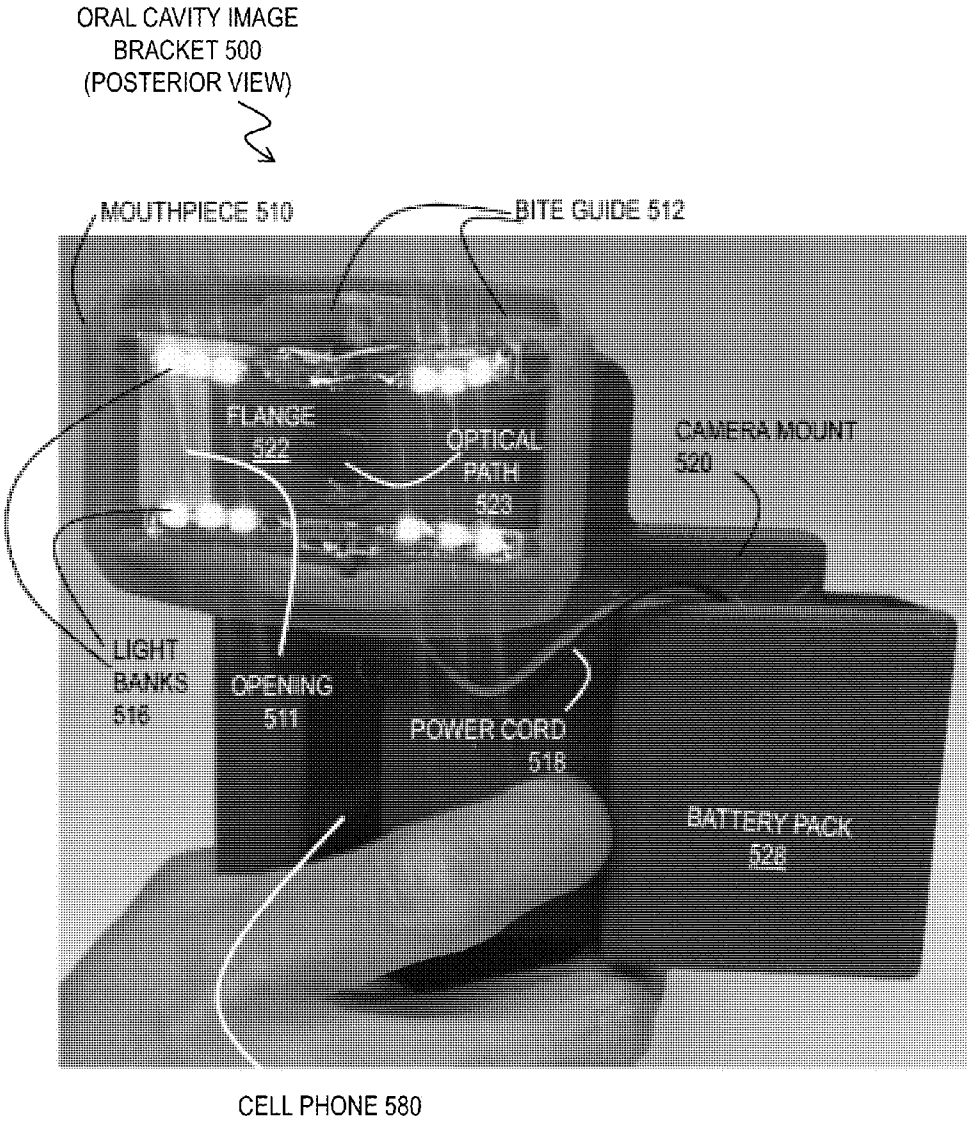

ORAL CAVITY IMAGE
BRACKET 500
(POSTERIOR VIEW)
MOUTHPIECE 510
BITE GUIDE 512
FLANGE
522
OPTICAL
PATH
528
CAMERA MOUNT
520
LIGHT
BANKS
516
OPENING
511
POWER CORD
518
BATTERY PACK
526
CELL PHONE 580

FIG. 5B

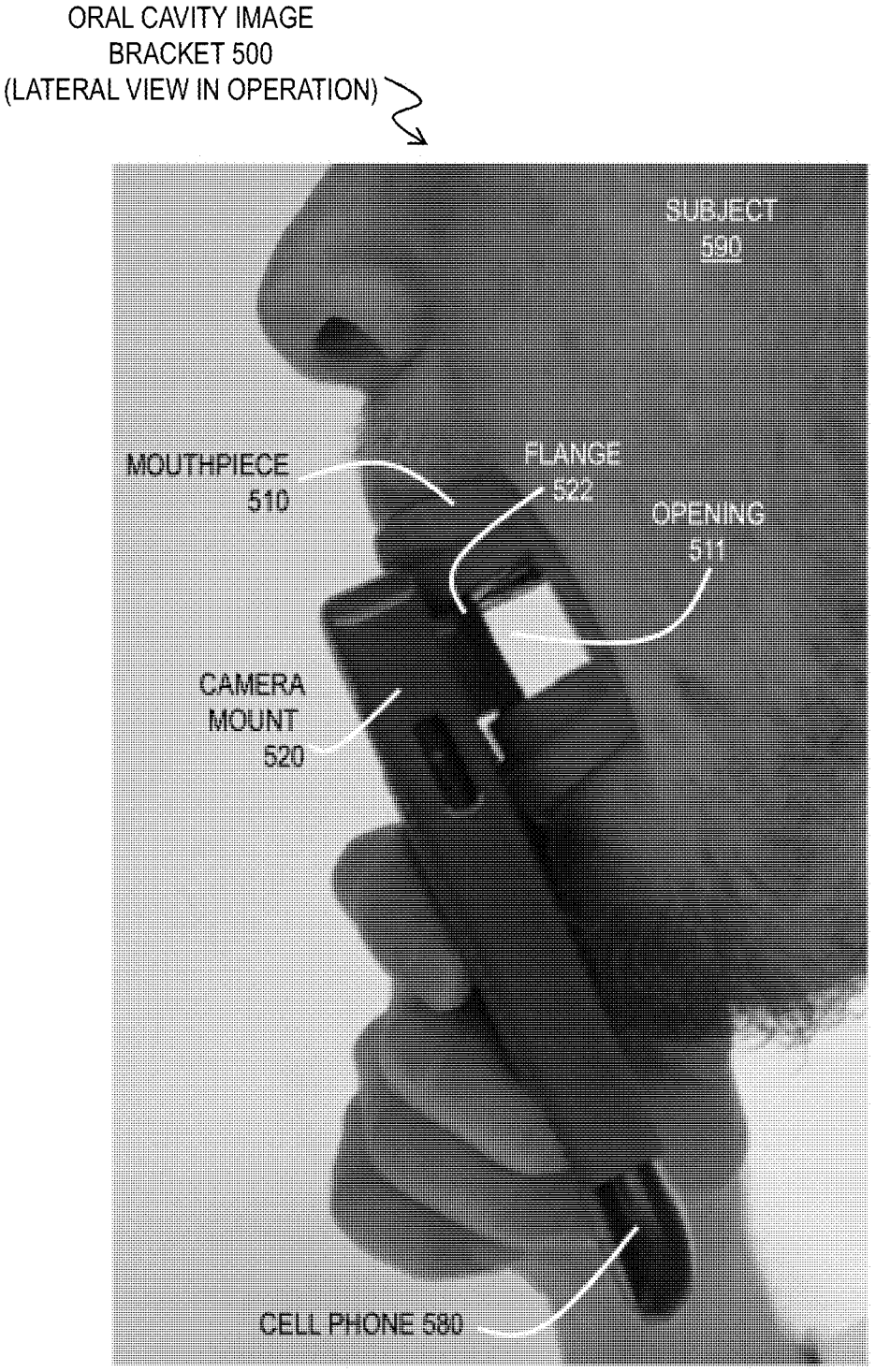
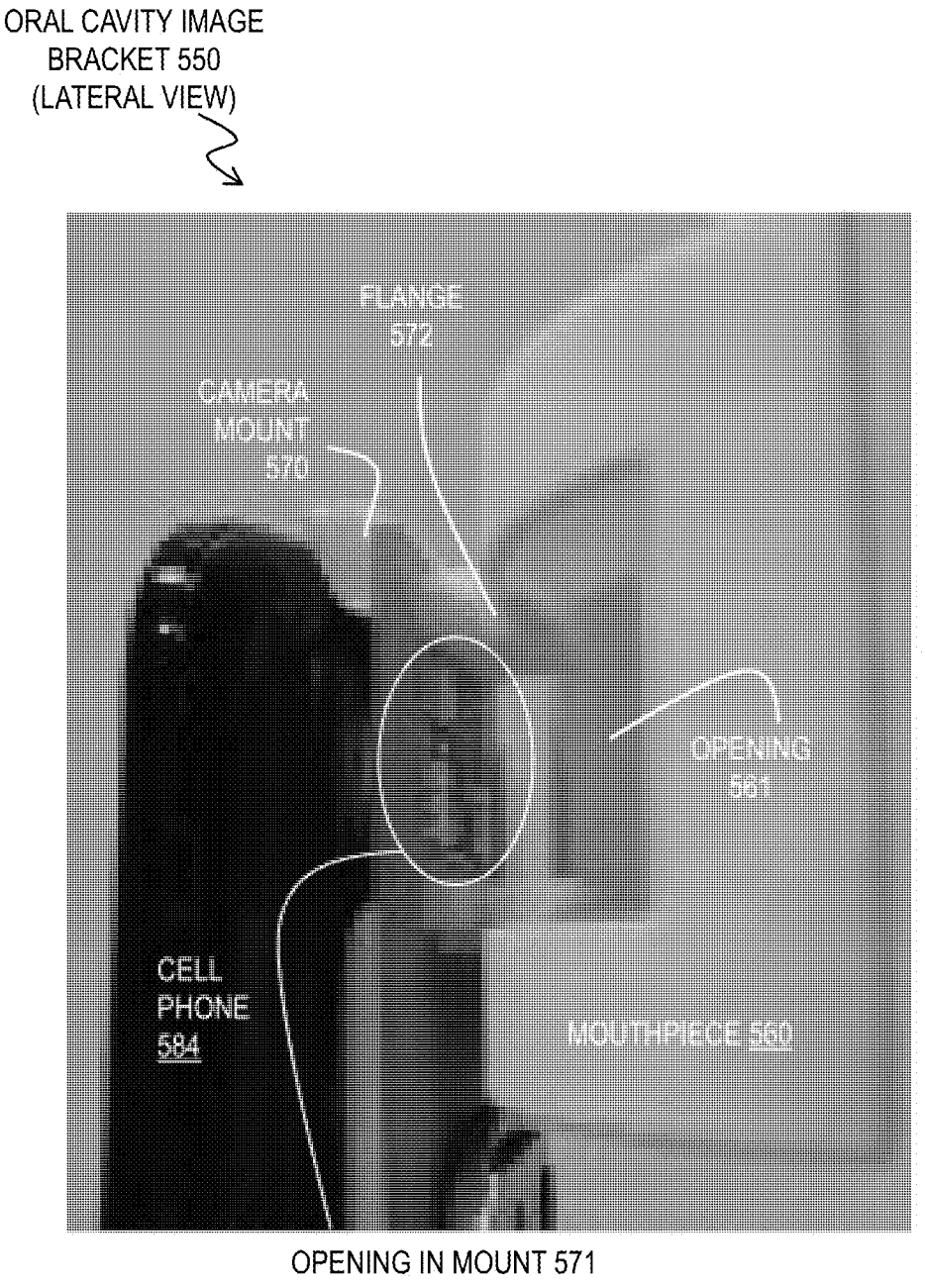

ORAL CAVITY IMAGE
BRACKET 500
(LATERAL VIEW IN OPERATION)
SUBJECT
530
MOUTHPIECE
510
FLANGE
522
OPENING
511
CAMERA
MOUNT
520
CELL PHONE 580

FIG. 5C



OPENING IN MOUNT 571

FIG. 5D

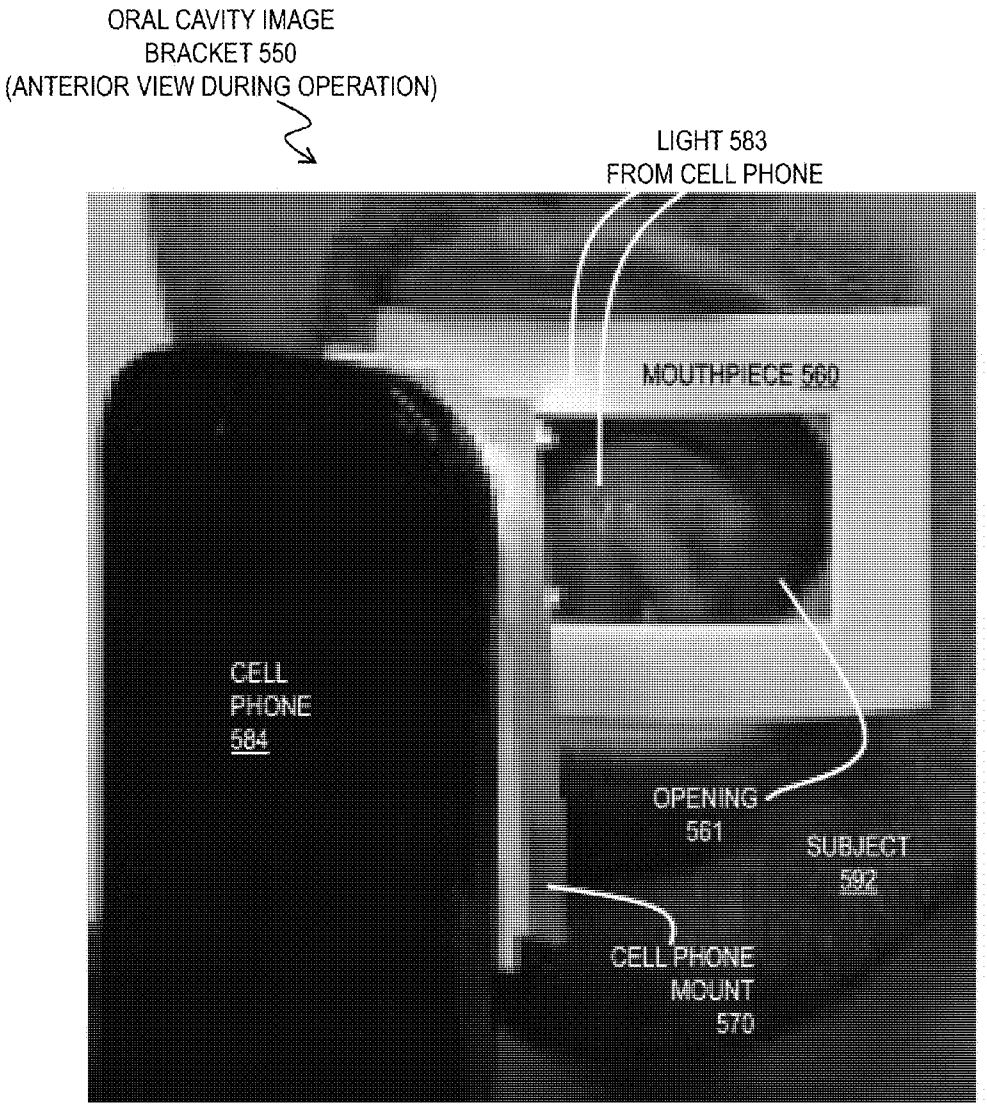

ORAL CAVITY IMAGE
BRACKET 550
(ANTERIOR VIEW DURING OPERATION)
LIGHT 583
FROM CELL PHONE
MOUTHPIECE 560
CELL
PHONE
584
OPENING
561
SUBJECT
592
CELL PHONE
MOUNT
570

FIG. 6A

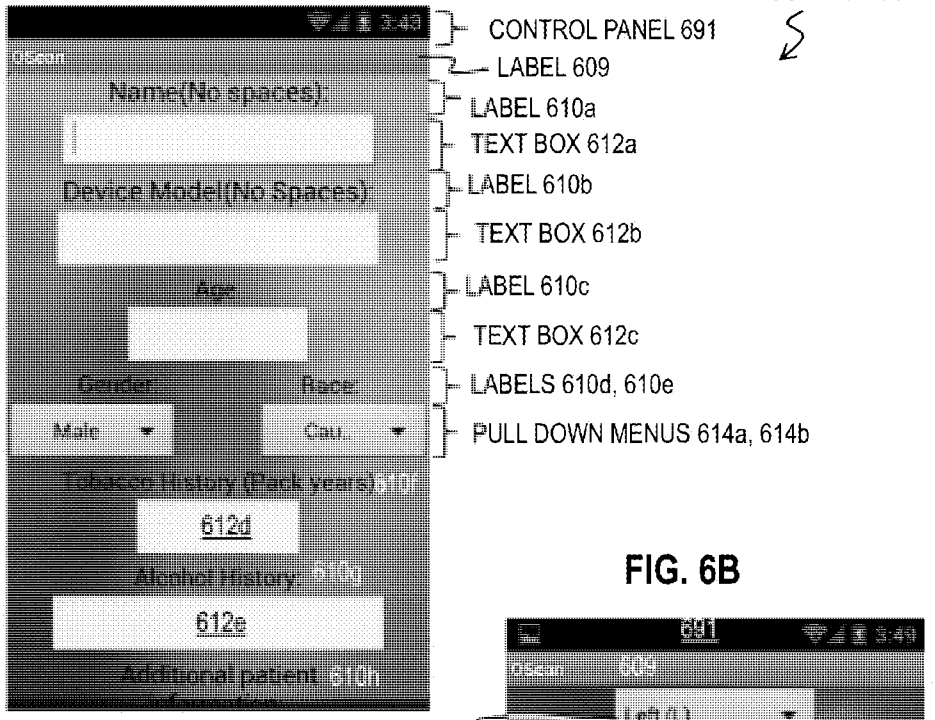

GUI PAGE 601
CONTROL PANEL 691
LABEL 609
LABEL 610a
TEXT BOX 612a
LABEL 610b
TEXT BOX 612b
LABEL 610c
TEXT BOX 612c
LABEL 610d, 610e
PULL DOWN MENUS 614a, 614b
Name(No spaces):
Device Model(No Spaces):
Age:
Gender
Race
Male
612d
612e

FIG. 6B

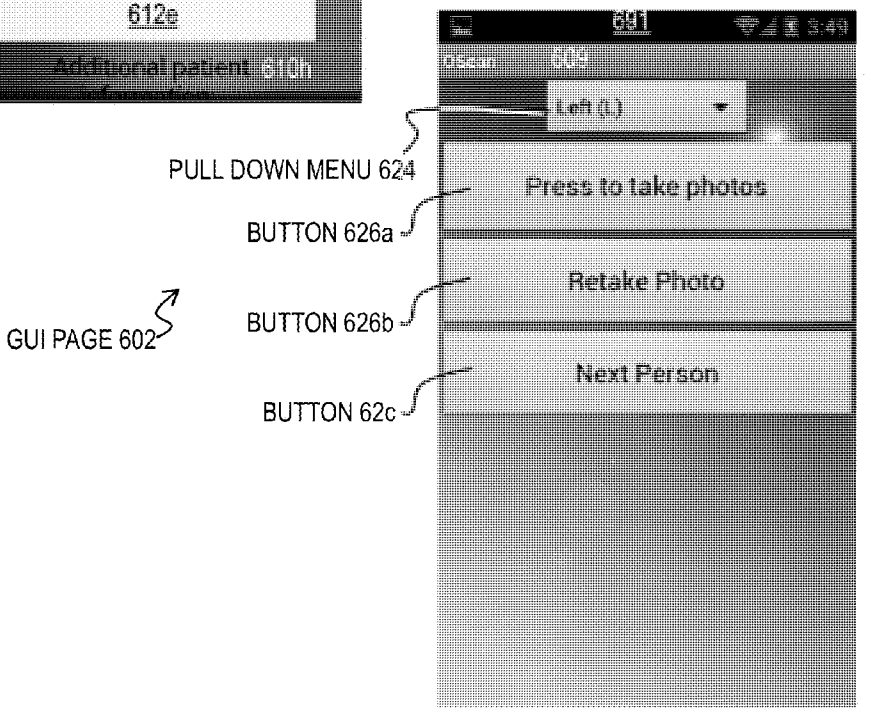

691
PULL DOWN MENU 624
BUTTON 626a
BUTTON 626b
BUTTON 62c
GUI PAGE 602
Left (L)
Press to take photos
Retake Photo
Next Person

FIG. 6C

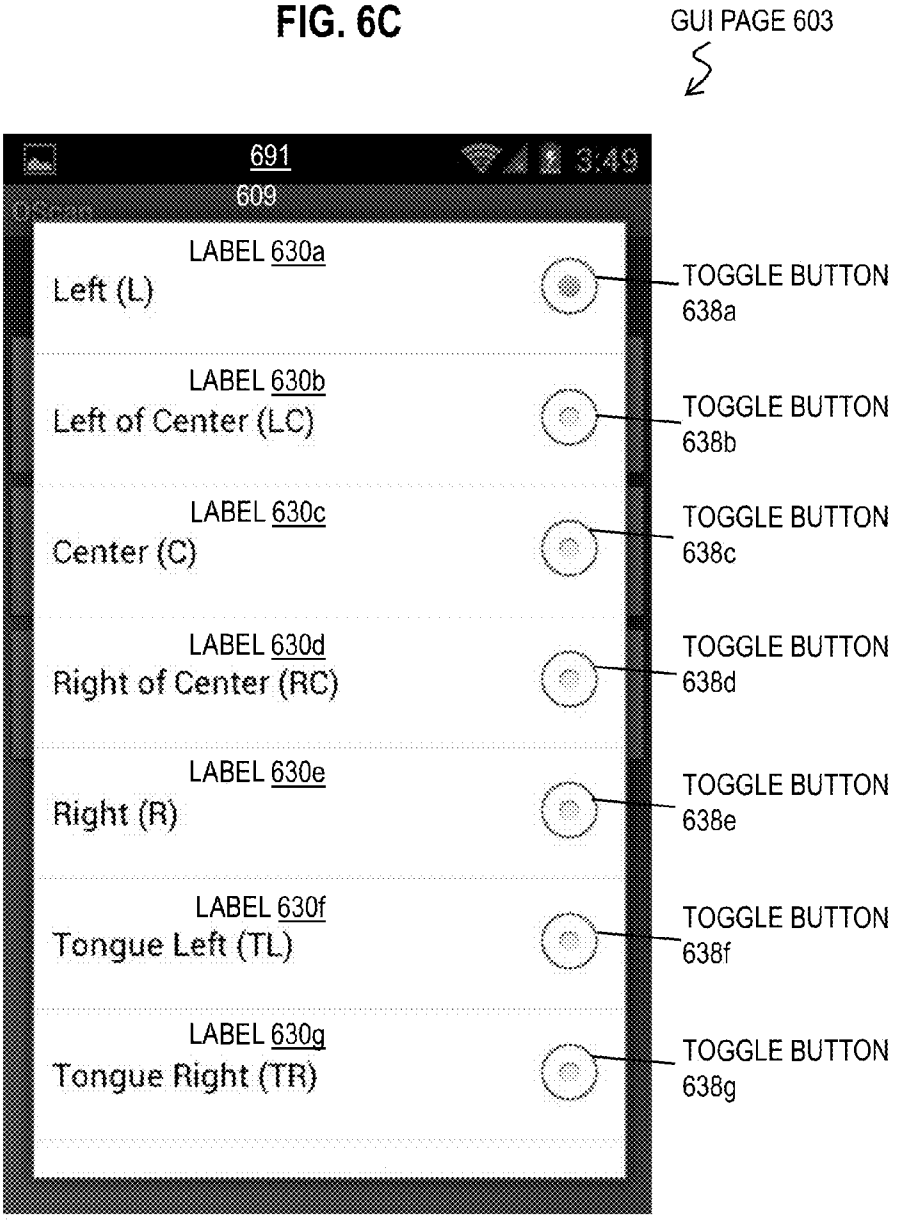

GUI PAGE 603
691
609
3:49
LABEL 630a
Left (L)
TOGGLE BUTTON 638a
LABEL 630b
Left of Center (LC)
TOGGLE BUTTON 638b
LABEL 630c
Center (C)
TOGGLE BUTTON 638c
LABEL 630d
Right of Center (RC)
TOGGLE BUTTON 638d
LABEL 630e
Right (R)
TOGGLE BUTTON 638e
LABEL 630f
Tongue Left (TL)
TOGGLE BUTTON 638f
LABEL 630g
Tongue Right (TR)
TOGGLE BUTTON 638g

FIG. 7

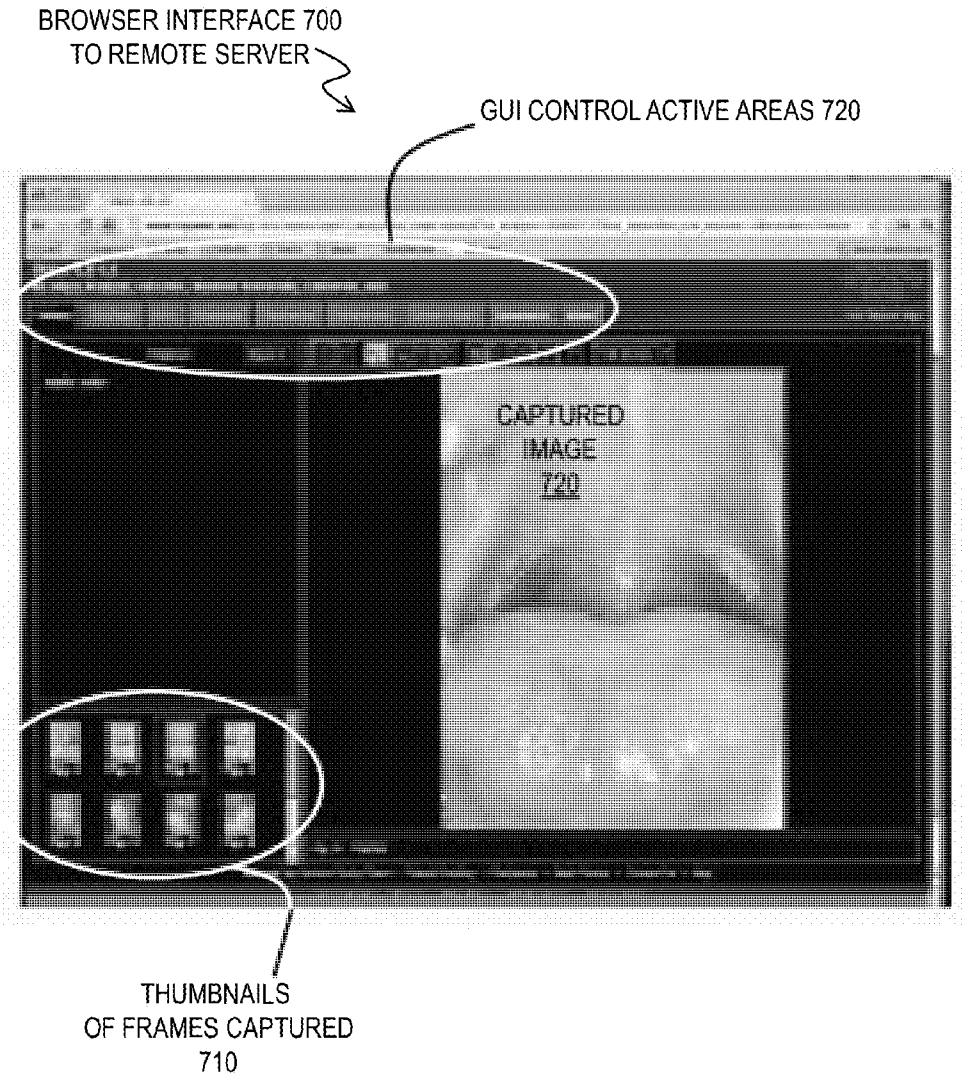

BROWSER INTERFACE 700
TO REMOTE SERVER
GUI CONTROL ACTIVE AREAS 720
CAPTURED
IMAGE
720
THUMBNAILS
OF FRAMES CAPTURED
710

FIG. 8

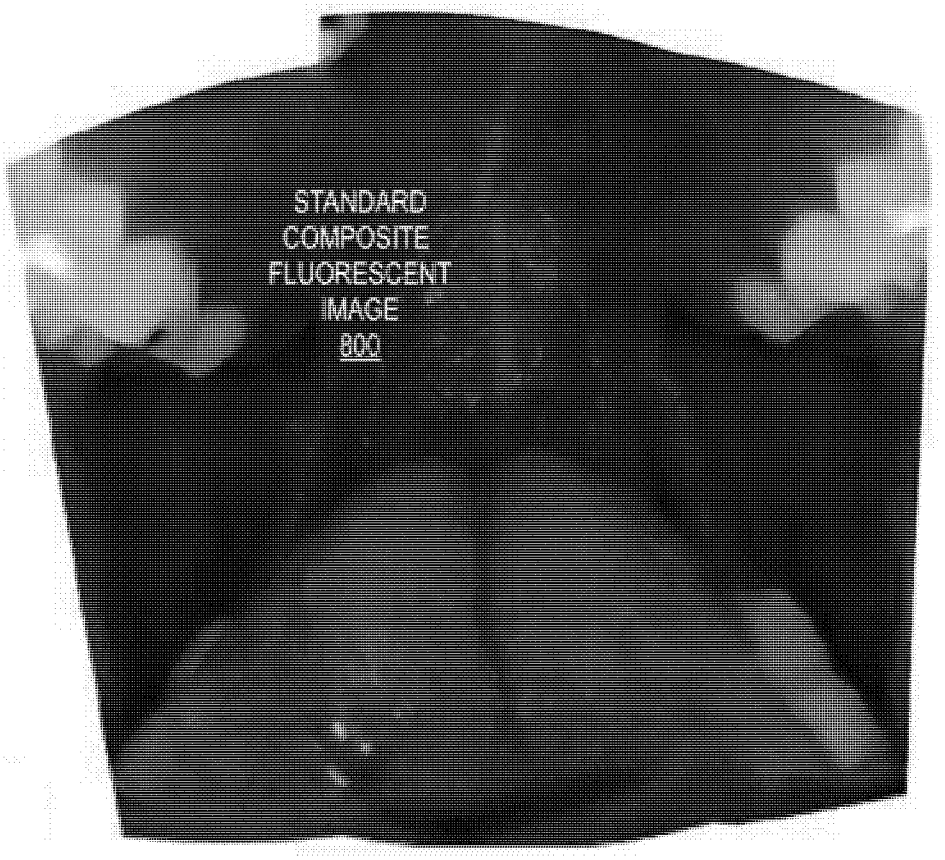


FIG. 9A FIG. 9B FIG. 9C

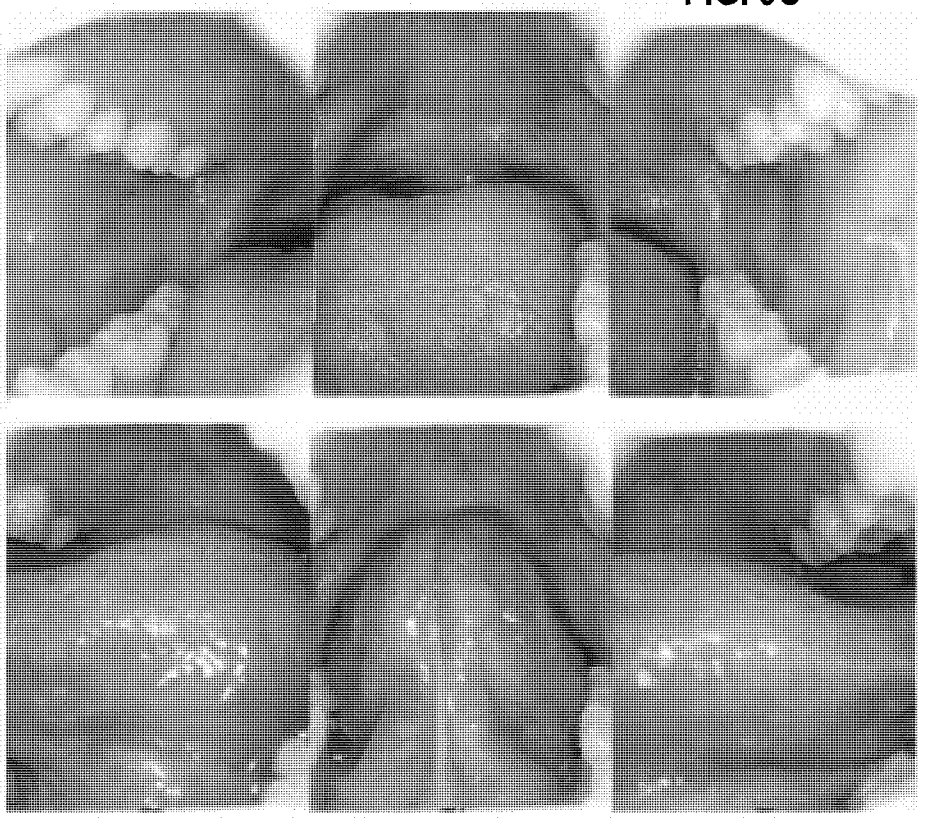


FIG. 9D FIG. 9E FIG. 9F

FIG. 10A

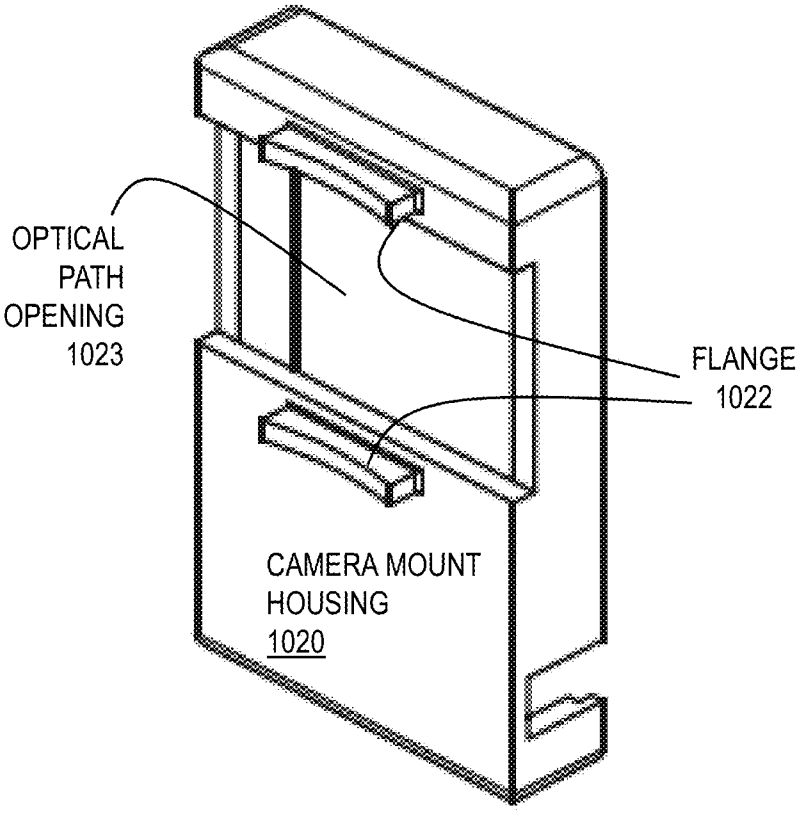


FIG. 10B

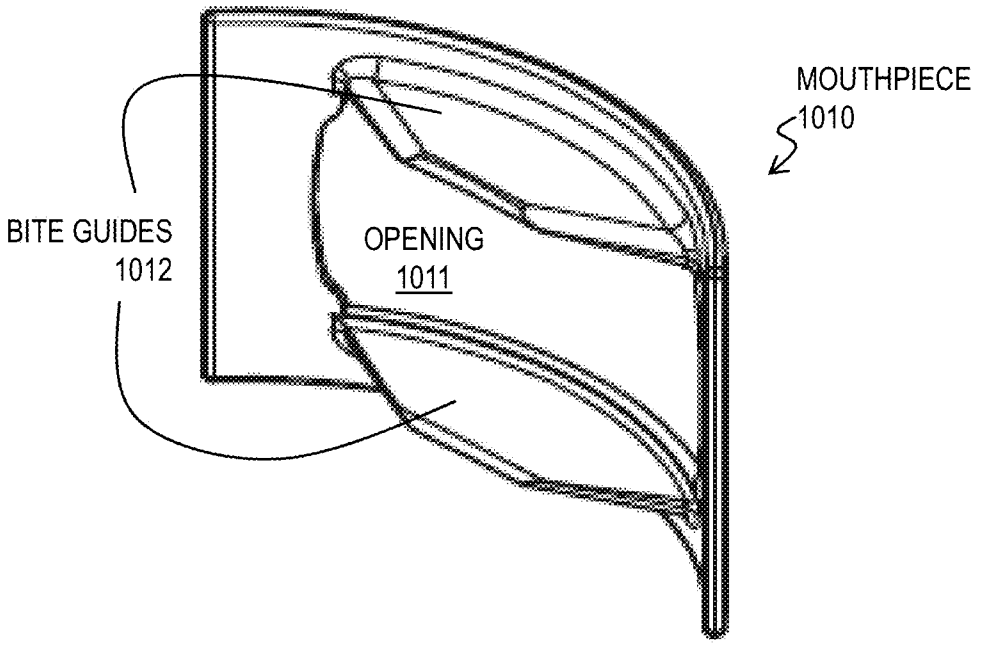


FIG. 10C

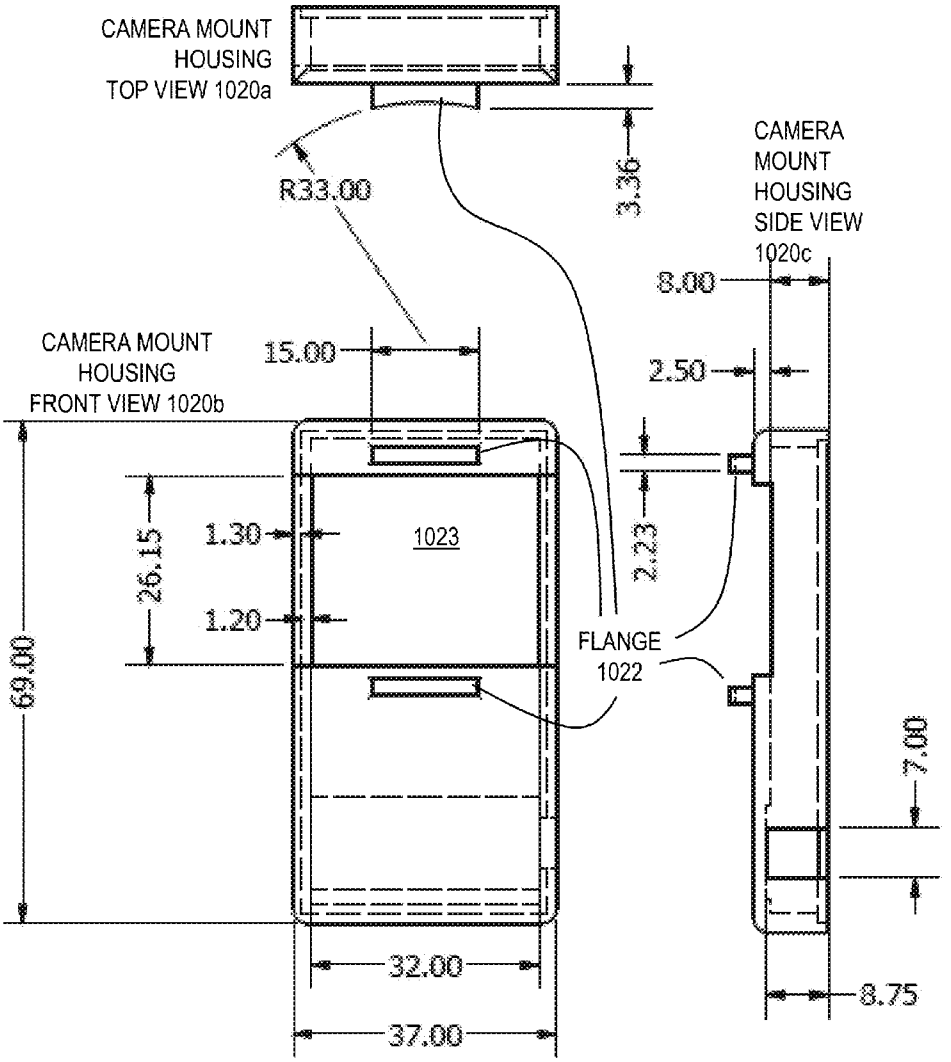
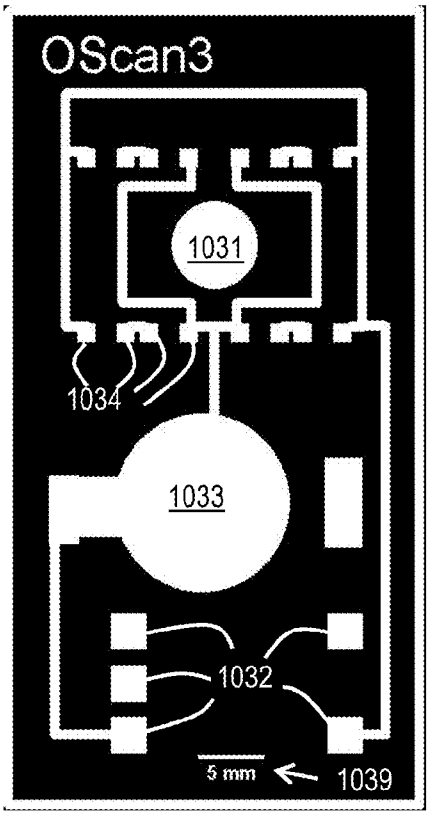

CAMERA MOUNT HOUSING TOP VIEW 1020a
R33.00
3.36
CAMERA MOUNT HOUSING SIDE VIEW 1020c
8.00
2.50
CAMERA MOUNT HOUSING FRONT VIEW 1020b
15.00
1.30
1.20
1023
26.15
69.00
2.23
FLANGE 1022
7.00
32.00
37.00
8.75

FIG. 10D



CAMERA MOUNT
CIRCUIT BOARD INSERT
1030

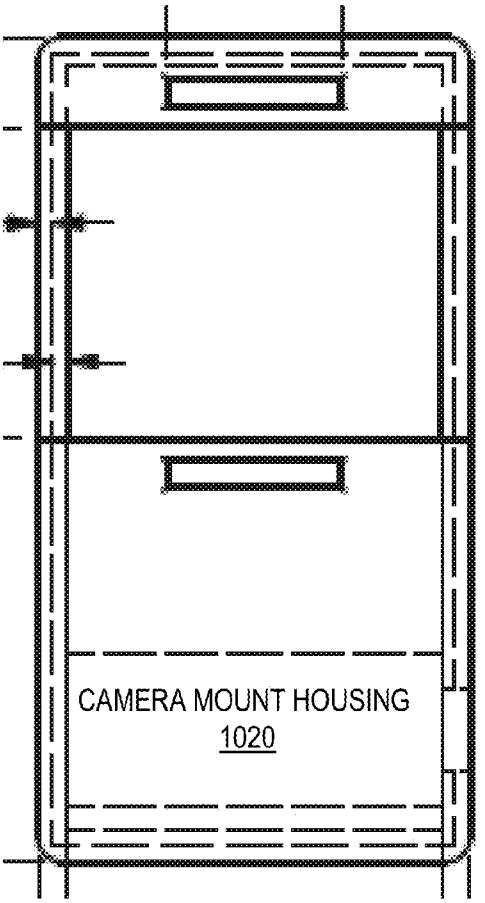


FIG. 10E

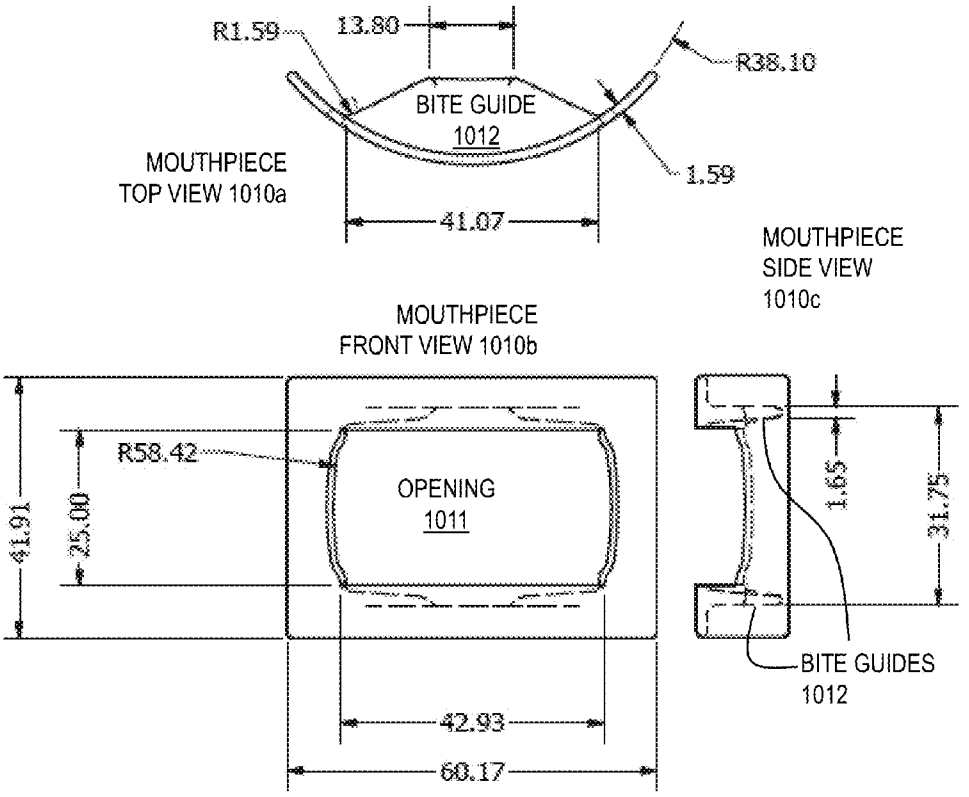


FIG. 10F

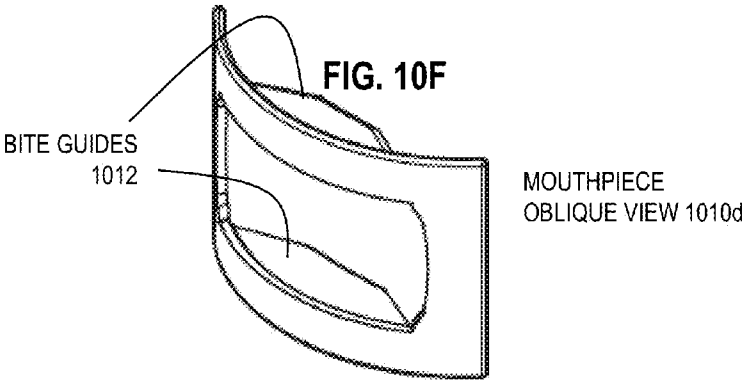


FIG. 11

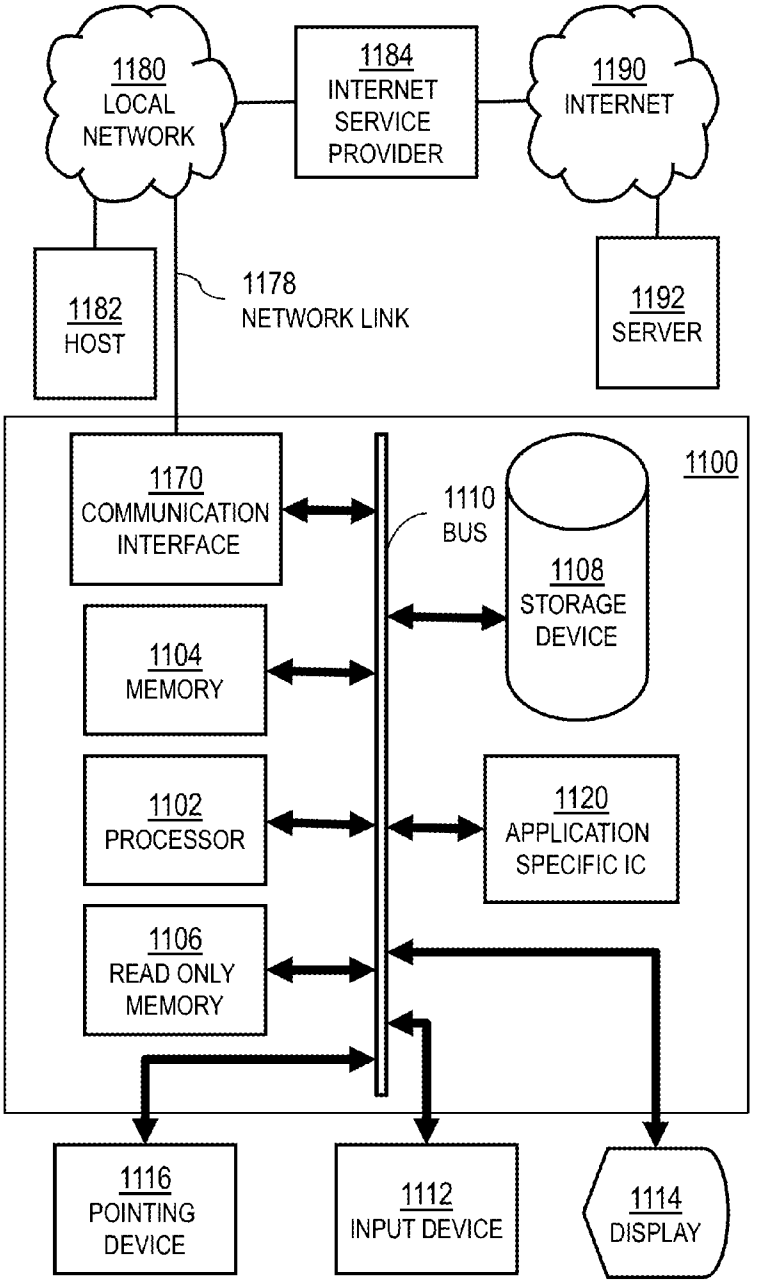
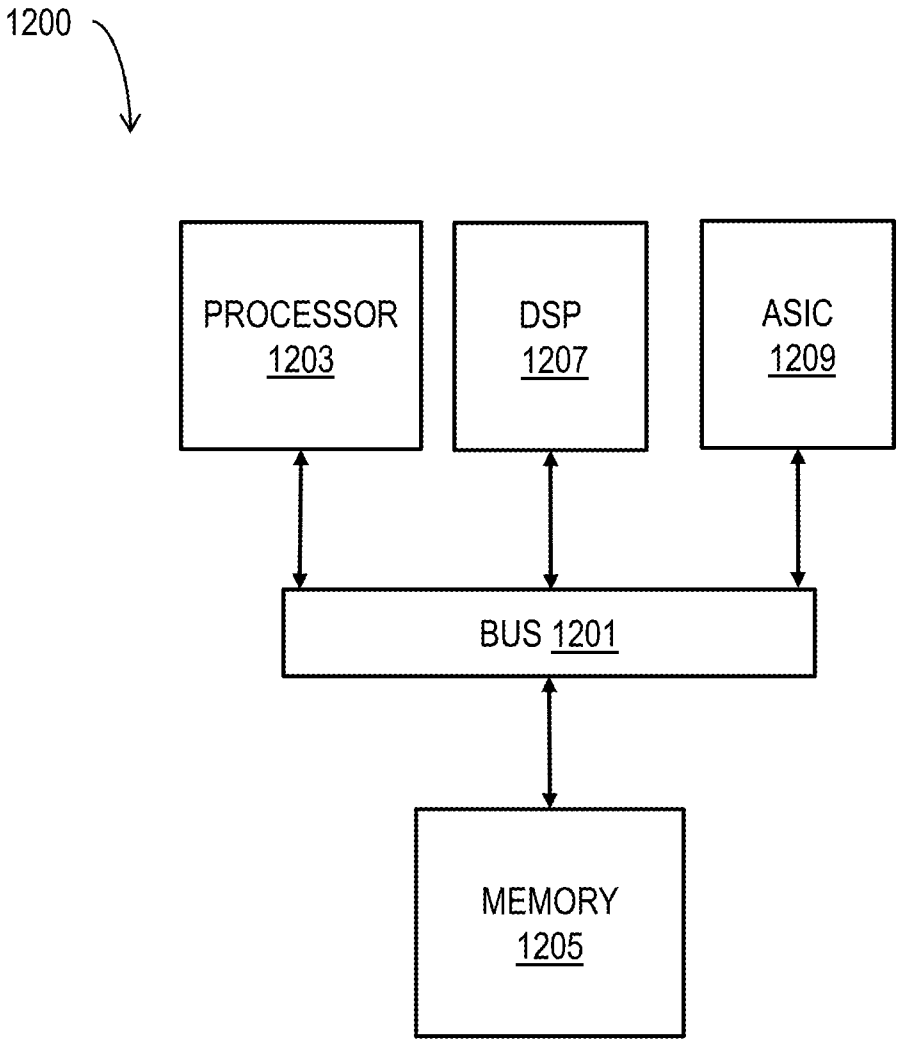

1180
LOCAL
NETWORK
1184
INTERNET
SERVICE
PROVIDER
1190
INTERNET
1182
HOST
1178
NETWORK LINK
1192
SERVER
1100
1170
COMMUNICATION
INTERFACE
1110
BUS
1108
STORAGE
DEVICE
1104
MEMORY
1102
PROCESSOR
1120
APPLICATION
SPECIFIC IC
1106
READ ONLY
MEMORY
1116
POINTING
DEVICE
1112
INPUT DEVICE
1114
DISPLAY

FIG. 12



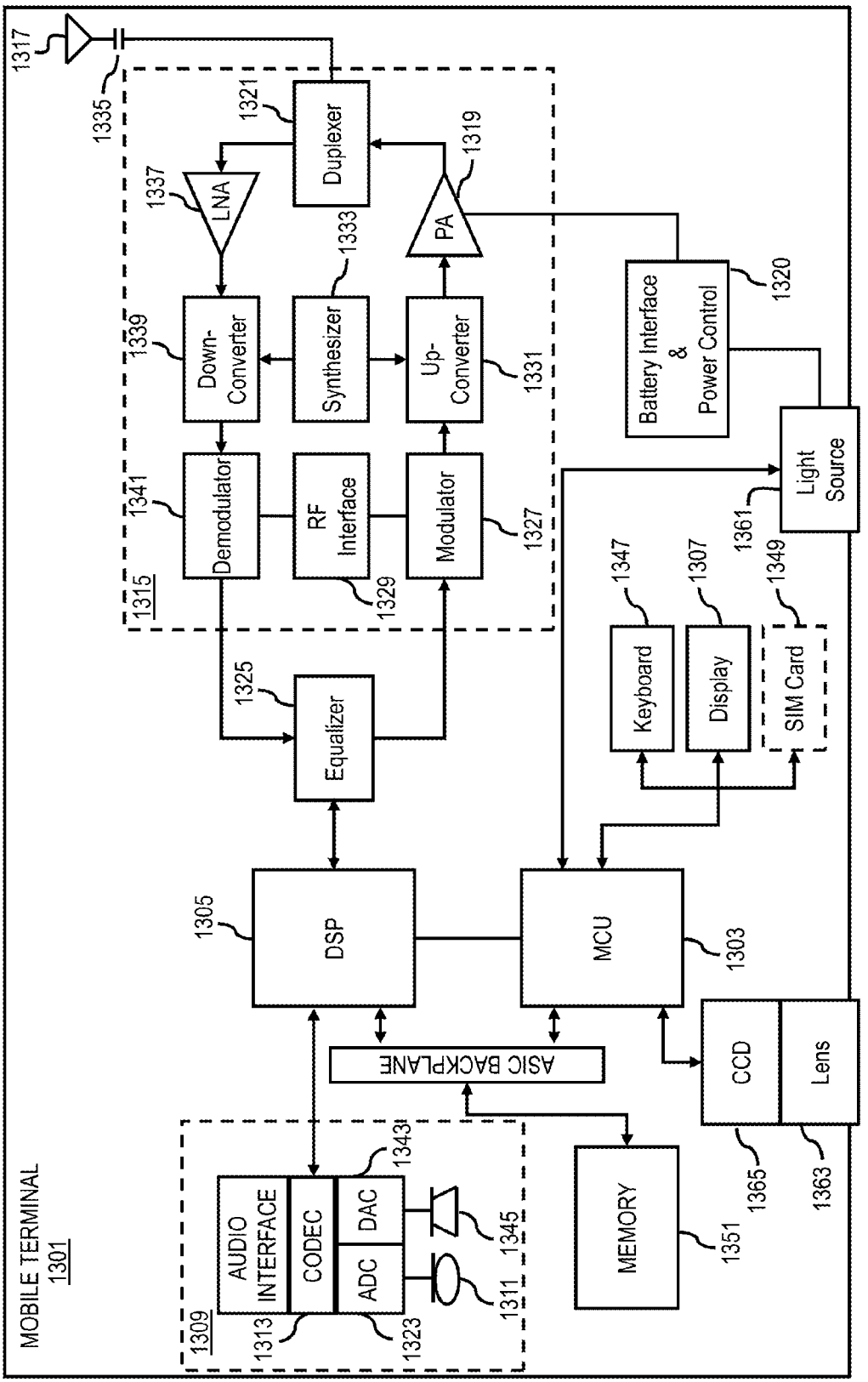

MOBILE TERMINAL
1301
1317
1335
1321
Duplexer
1319
LNA
1337
PA
1333
Down-Converter
1339
Synthesizer
Up-Converter
1331
Battery Interface & Power Control
1320
Demodulator
1341
RF Interface
Modulator
1327
1315
1329
Light Source
1361
1347
1307
1349
Keyboard
Display
SIM Card
1325
Equalizer
1305
DSP
MCU
1303
ASIC BACKPLANE
CCD
Lens
1309
AUDIO INTERFACE
CODEC
DAC
ADC
1343
1345
1311
1313
1323
MEMORY
1351
1365
1363

FIG. 13

TECHNIQUES FOR STANDARDIZED IMAGING OF ORAL CAVITY

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims benefit of Provisional Appln. 61/597,772, filed Feb. 11, 2012, the entire contents of which are hereby incorporated by reference as if fully set forth herein, under 35 U.S.C. §119(e).

BACKGROUND OF THE INVENTION

[0002] The oral cavity in humans is an indicator of a number of diseases, including submucus fibrosis, gingivitis, oral cancer and others. For example, 400,000 cases of oral cancer are reported world-wide, one third of which are reported in developing countries with less than one doctor for every 50,000 patients. Tobacco chewing and smoking is the primary driver for the same where 70% of world tobacco consumption is in developing countries. Late discovery of an oral cancer has survival rates of 50% while early stage detection can improve the rate to more than 90%.

SUMMARY OF THE INVENTION

[0003] Even though the oral cavity is widely accessible for examination, no standards exist for comprehensive imaging of the cavity. Techniques are provided for some combination of inexpensive, efficient, comprehensive or standardized imaging of the oral cavity.

[0004] In a first set of embodiments, an oral cavity image bracket includes a mouthpiece and a camera mount. The mouthpiece of rigid material includes an upper bite guide and a lower bite guide, both disposed on a posterior side of the mouthpiece and separated by an opening through the mouthpiece. The upper bite guide and lower bite guide are spaced apart such that a subject biting down on the upper bite guide with the subject's upper jaw and biting up on the lower bite guide with the subject's lower jaw opens the subject's oral cavity to inspection through the opening. The camera mount is disposed on an anterior side of the mouthpiece and includes a flange configured to engage and slide along the opening of the mouthpiece. The camera mount further comprises an optical path and a clip. The optical path is configured for light to pass through the camera mount and through the opening in the mouthpiece. The clip is disposed on an anterior side of the camera mount, and is configured to removeably hold a camera on the anterior side of the camera mount to record light passing through the optical path from the posterior side of the camera mount.

[0005] In some embodiments of the first set, the camera is a cell phone with built in camera and on board processor.

[0006] In some embodiments of the first set, the mouthpiece further comprises a light source disposed on the posterior side of the mouthpiece and configured to illuminate the oral cavity of the subject.

[0007] In some embodiments of the first set, the optical path comprises a removeable optical filter that blocks light from a light source and passes fluorescent light emitted by tissue in the oral cavity of the subject in response to the light source.

[0008] In a second set of embodiments, an oral cavity imaging system includes the oral cavity image bracket described above plus a camera and a processor. The camera is removeably attached to the anterior side of the camera mount, and is configured to record and display an image based on light passing through the optical path from the posterior side of the camera mount. The processor is configured to merge data from a plurality of images recorded by the camera at a corresponding plurality of positions of the camera mount as the camera mount slides along the opening in the mouthpiece. In some embodiments of this set, the camera is a cell phone with built in digital camera and the processor on board the cell phone.

[0009] In a third set of embodiments, a method includes removeably attaching a camera to a camera mount of the oral cavity image bracket comprising the camera mount and a mouthpiece described above. The method also includes causing a subject to bite against the bite guides of the mouthpiece. The method further includes sliding the camera mount to a plurality of positions along the opening in the mouthpiece and causing the camera to capture a plurality of images corresponding to the plurality of positions. In some embodiments of this set, the method further comprises using a processor to merge data from the plurality of images into a standard image.

[0010] In a fourth set of embodiments, a computer program method includes determining distance from an imaging plane to a surface of an oral cavity of a subject based on relative intensity of a pixel in an image frame captured at the imaging plane compared to adjacent pixels in the image frame and a model of focusing optics. This is performed for each of multiple image frames of the oral cavity corresponding to multiple different look directions. The method includes merging, into a single image, pixels from the multiple image frames of the oral cavity.

[0011] In other sets of embodiments, an apparatus or computer readable medium is configured to cause the apparatus to perform one or more steps of the computer program method.

[0012] Still other aspects, features, and advantages of the invention are readily apparent from the following detailed description, simply by illustrating a number of particular embodiments and implementations, including the best mode contemplated for carrying out the invention. The invention is also capable of other and different embodiments, and its several details can be modified in various obvious respects, all without departing from the spirit and scope of the invention. Accordingly, the drawings and description are to be regarded as illustrative in nature, and not as restrictive.

BRIEF DESCRIPTION OF THE DRAWINGS

[0013] The present invention is illustrated by way of example, and not by way of limitation, in the figures of the accompanying drawings and in which like reference numerals refer to similar elements and in which:

[0014] FIG. **1** is a block diagram that illustrates an expensive apparatus for imaging the oral cavity, as in the prior art;

[0015] FIG. **2**A, FIG. **2**B and FIG. **2**C are block diagrams that illustrate an example oral cavity image bracket and system, according to an embodiment;

[0016] FIG. **3**A, FIG. **3**B and FIG. **3**C are block diagrams that illustrate example fields of view from multiple positions of a camera mount along an opening in a mouthpiece of the example oral cavity image bracket, according to an embodiment;

[0017] FIG. **3**D is a block diagram that illustrates example overlapping image frames captured from the multiple positions of the camera mount, according to an embodiment;

[0018] FIG. **4**A is a flow chart that illustrates an example method for producing and using one or more standard images

of the oral cavity utilizing the oral cavity image bracket and a cell phone with built in camera and processor, according to an embodiment;

[0019] FIG. **4**B is a flow chart that illustrates an example computer program method for producing a standard image of the oral cavity based on overlapping image frames captured from the multiple positions of the camera mount, according to an embodiment;

[0020] FIG. **5**A and FIG. **5**B are photographs that illustrate an example oral cavity image bracket and system, according to another embodiment;

[0021] FIG. **5**C and FIG. **5**D are photographs that illustrate an example oral cavity image bracket and system, according to yet another embodiment;

[0022] FIG. **6**A through FIG. **6**C are block diagrams that illustrate an example graphical user interface (GUI) for an oral cavity image application for a programmable cell phone with built-in camera and processor, according to another embodiment;

[0023] FIG. **7** is a block diagrams that illustrate an example browser graphical user interface (GUI) for a remote server configured for processing oral cavity images, according to another embodiment;

[0024] FIG. **8** is an image that illustrates an example standardized fluorescence image for an oral cavity, according to an embodiment;

[0025] FIG. **9**A through FIG. **9**F are images that illustrate example bright field images for an oral cavity, according to an embodiment;

[0026] FIG. **10**A through FIG. **10**F are scaled drawings for a mouthpiece and camera mount, according to a particular embodiment;

[0027] FIG. **11** is a block diagram that illustrates a computer system upon which an embodiment of the invention may be implemented;

[0028] FIG. **12** illustrates a chip set upon which an embodiment of the invention may be implemented; and

[0029] FIG. **13** is a block diagram that illustrates example components of a mobile terminal (e.g., a cell phone handset) for communications, which is capable of operating in the system of FIG. **2**C, according to one embodiment.

DETAILED DESCRIPTION

[0030] A method and apparatus are described for standardized imaging of the oral cavity. In the following description, for the purposes of explanation, numerous specific details are set forth in order to provide a thorough understanding of the present invention. It will be apparent, however, to one skilled in the art that the present invention may be practiced without these specific details. In other instances, well-known structures and devices are shown in block diagram form in order to avoid unnecessarily obscuring the present invention.

[0031] Some embodiments of the invention are described below in the context of using a smart phone, e.g., a programmable cell phone with on board (built-in) processor and digital camera. However, the invention is not limited to this context. In other embodiments, other cameras are used, digital or otherwise, with or without on board processors. For example, in embodiments using a digital camera without a programmable on board processor, a separate programmable processor is provided on the mouthpiece or camera mount, and digital output from the camera is fed into the separate processor.

1. Overview

[0032] In many countries of the world, there is a lack of adequate dental screening and care. For example, in India, the Primary Health Centre (PHC) is the first point of contact for the community to get health care service and has no set criteria for employment of dentists in the rural areas of the country, leading to random appointments across the country. This has resulted in a huge shortfall in the oral care services at the primary level with private services being very cost-intensive.

[0033] Though a person's oral cavity is easily accessible via a regular inspection, it has been a challenge for clinicians and health-workers alike to assess the state of oral health of an individual in resource poor settings. The reasons for the same include complete lack of trained health workers who can assess oral hygiene problems purely by an inspection. Even in hospital settings at field sites without expensive equipment, it is difficult for clinicians to distinguish subtle neoplastic changes in oral mucosa from other inflammatory diseases.

[0034] FIG. **1** is a block diagram that illustrates an expensive apparatus for imaging the oral cavity, as in the prior art. The expensive apparatus is a head mount surgical light **100** that is placed on the head of a clinician in order to inspect and photograph the oral cavity of a subject **190**, such as a patient. Mounted to the front of the head mount **110** is a light source **114** that illuminates the oral cavity when the mouth of the subject **190** is opened. The light emitted in two directions is captured in the two receptors of the binocular loupes **116** with optical filters. The two images are collected by the compact color camera **118**, which is often a digital camera. Power for the light source **114** and camera **118** is provided by a lithium ion battery **112** worn by the clinician in a pocket of clothing or attached to a belt or other harness. For fluorescence imaging useful for some diagnoses, the optical filters of the binocular loupes **116** pass light in the fluorescence emission wavelength and block light in the other wavelengths of the illuminating light source **114**. Such delicate and expensive equipment as head mount surgical light **100**, including a trained operator, is in rare supply and often completely lacking in the rural areas of many countries.

[0035] The complete lack of screening and diagnostics tools for oral cancer in developing countries, where two-third of the world's cases of oral cancer exist and 70% of world consumption of tobacco occurs, clearly marks a very large gap in the current infrastructure (commercial and social) for tackling oral cancer worldwide.

[0036] Current methodology to document lesions found in oral cancer patients are based on manual sketches and are insufficient for conveying quantitative information to either the clinician or the patient. Such sketches used to document the progress of a lesion must be relied upon by radiotherapists to determine if a therapy is successful on many rural patients. A standardized oral cavity scanner would allow for a quantitative method for clinicians to track the progress of the lesions in individual patients before and during therapy.

[0037] Inventors have identified several needs that arise from discussions with current practicing field oncologists, oral and neck surgeons and cancer screening programs. These needs include the critical need for an ultra-wide angle imaging of the oral cavity, and image standardization for comparative analysis between different scans of the same patient and among different patients. The wide angle and standard images allow clinicians to assess the damage, and assess recovery during treatment. Furthermore, integration of auto-

fluorescence imaging in such a platform allows the clinician to assess the epithelium in the oral cavity for early markers of oral cancer. Auto-fluorescence imaging of the oral cavity for early markers of oral cancer include illumination with light in a first wavelength band and detection of light in a different auto-fluorescence wavelength band, as described in more detail below.

[0038] However, the real technical difficulty in comprehensive imaging of the oral cavity can be experienced simply by trying to take some images of the oral cavity using readily available commercial cameras, such as camera-phones like an IPHONE™ of APPLE INC.™ of Cupertino, Calif.

[0039] Inventors have developed a low cost oral cavity scanning tool based, in some embodiments, on widely available camera-phones. Two key technical innovations pioneered by the inventors make fast and comprehensive fluorescence imaging of the oral cavity possible using commercially available camera-phones in a matter of seconds. This ease of use with standardization of imaging angles (as is commonly used in X-ray imaging) opens the door for rapid diagnostics and classification of oral inflammations from malignant oral cancer. Advantageous aspects originate from geometrical optics applied to the specific problem of imaging the surface of the oral cavity.

[0040] In some embodiments, pinhole based camera-phone imaging is employed. The challenge of ultra-wide angle imaging of a spherical cavity can be met by an unconventional complete imaging technique known as pinhole imaging that does not rely on a material lens. Rather than using a focusing lens to project an image on a plane coincident with a sensor plane (e.g., a charge coupled device, CCD, array), a miniature pinhole is used to keep out-of-focus light from impinging on the sensor plane. This technique also eliminates the use of costly optical components that are difficult to mass-produce due to high manufacturing tolerance required for lenses made at small length scale. A small hole (with diameters in a range from about 1 micrometers, μm, to about 1000 μm, depending on ambient conditions, where 1 $\mu m = 10^{-6}$ meters) can project light from the oral cavity directly onto a screen/wall/surface or be coupled to a camera/camera phone. For a pinhole imaging system, in some embodiments, a simple screen is used to form a real image and the camera phone or other camera takes an image of the screen (which is a real image) using native camera phone lenses and optics. In other embodiments, a direct capture method is used in which a bare CCD array is utilized (e.g., by removing the native lenses of a camera phone), which captures a direct image that forms on the bare CCD. Pinhole imaging provides an advantage of a very large field of view image that is in focus everywhere across the image.

[0041] In some embodiments, modular optical components, including a fish-eye lens in some embodiments, are used with optical transforms. To allow for 360 degree imaging of the oral cavity, the inventors have exploited the use of fish-eye lenses that snap onto any regular cellphone and thus provide a unique and complete coverage of the oral cavity. Since the geometrical transformation of the object and the image are both known, simple optical transforms in post-processing reverses the image to correct for optical aberration and provides a true area of lesions for diagnostic purposes.

[0042] Further, in some embodiments, the inventors have integrated an auto-fluorescence imaging scheme, with excitation wavelength of 488 nanometers (nm, 1 $nm = 10^{-9}$ meters) provided by a blue light emitting diode (LED), and an optical filter deployed to block the excitation wavelength and other wavelengths, or pass just the fluorescence wavelengths between about 500 nm to about 600 nm (in the green range), depending on illness, and a cell phone camera acting as the imaging sensor.

[0043] Furthermore, in some embodiments, software tools are specifically written for post-processing, segmentation and labeling of these images to categorize collected samples in various bins based on several key morphological indicators in the given images. In various embodiments, a clustering approach is either trained to provide an autonomous diagnostic or is used as an aid for a trained specialist who finally labels the images as oral inflammations or suspected lesions likely to be oral cancer. In most cases, diseased tissue causes a region that is dimmer than the surrounding tissue. The exact percentage dimmer is determined in clinical trials. This threshold varies the clinical sensitivity of the device. In some embodiments, a dye (such as a Toluidine blue stain) is applied to the oral cavity of the subject before the image frame data is collected, to enhance the distinction between healthy and diseased tissue in the fluorescence frames.

[0044] It is anticipated that some embodiments will include a rapid screening platform, which allows for regular and large scale (hundreds of patients per worker per day) screening, and which is followed by traditional histological pathology based detection of cancer type in suspected cases. Rather than reinvent the wheel; it is anticipated further to use existing open-source mobile-programming platforms for healthcare delivery in the hardware and software tools, in various embodiments. This allows for an extremely critical, user-friendly and culturally acceptable graphical user interface (GUI). This also allows for rapid launch and test of the devices and methods described herein.

[0045] Plug-and-play modular optical components mounted on traditional camera-phones specifically target comprehensive and standardized imaging of the oral cavity. Pinhole based imaging is applied to camera-phones in some embodiments for ultra-low cost and robust imaging platforms. The GUI of such camera phones provide ease of use without prior training to make imaging the oral cavity as simple and intuitive as taking picture of a person's face. Standardization of imaging by mechanical constraints, applied using a simple mouthpiece on which the patient bites, automatically positions the imaging plane (camera) at a specific relative coordinate to the patients face.

[0046] This easy accessibility combined with convincing individuals of the great importance of oral hygiene in India (and similar low to middle income countries) presents itself as an opportunity to develop a health care delivery system based on a comprehensive, low-cost and fast imaging tool for the oral cavity.

[0047] FIG. **2**A, FIG. **2**B and FIG. **2**C are block diagrams that illustrate an example oral cavity image bracket **200** and system **250**, according to an embodiment. In FIG. **2**A a posterior view of the mouthpiece from a clinician's point of view is depicted. The posterior side of the bracket **200** is presented to a subject, such as a patient, to bite upon.

[0048] A mouthpiece **210** includes an upper bite guide **212***a* and a lower bite guide **212***b*, collectively referenced hereinafter as bite guides **212**. The gums or teeth of the subject's upper jaw are placed inside the upper bite guide **212***a*. As depicted, the upper bite guide **212***a* comprises a flat area jutting out of the plane of the view of FIG. **2**A far enough to accommodate the teeth or gums of a subject's upper jaw

and at the outer edge turning upward to form the depicted trapezoidal shape. Similarly, the lower bite guide **212***b* comprises a flat area jutting out of the plane of the view of FIG. **2**A far enough to accommodate the teeth or gums of a subject's lower jaw and at the outer edge turning downward to form the depicted trapezoidal shape. In various other embodiments, one or both of the bite guides **212** occupy a larger or smaller portion of the left to right dimension of the mouthpiece or involve a horizontal outward jutting portion with a groove configured to accommodate the gums or teeth of a subject, or some combination.

[0049] The mouthpiece **210** is made of a rigid material that does not substantively change shape when subjected to the pressures of the bite of a subject (e.g., changes shape less than about 20% and preferably less than about 1%). For example, in various embodiments the mouthpiece material is one or more of Nylon, Silicone, ABS plastic, Medical grade ABS plastic, polyurathane (mercury free), PMMA, PDMS, PET, Acrylic, Acetal Copolymet, Acetal Homopolymer, LCP, LDPE, LLDPE, Polycarbinate, PBT, PETG, polypropylene, thermoplastic elastomers, (basically any plastic or rubber that will not harm the human body), high density rigid foam, and medium density rigid foam. In various other embodiments, multiple materials are used in combination where one or more disposable parts are made from one kind of plastic (such as one listed above or otherwise) and the second portion is built from another plastic (such as one listed above or otherwise).

[0050] The bite guides are separated by an opening **211** that provides a view into the subject's oral cavity. As depicted, the opening **211** is closed at both left and right sides. However, in some embodiments the opening extends to the left side of the mouthpiece or to the right side of the mouthpiece. As depicted, the mouthpiece **210** is an integral piece of material; however, in some embodiments the left or right side of the opening **211** is a post made of a different material.

[0051] In various embodiments, the dimensions of the mouthpiece **210**, bite guides **212**, and opening **211** are configured to fit the mouths of patients of a certain size range, such as adults, adolescents, children or infants. For example, in some embodiments the mouthpiece dimensions are selected from a group of standard sizes, from about 1.5 centimeters to about 6 centimeters. The larger size the subject can withstand, the more of the subject's mouth can be imaged. Thus, the upper bite guide **212***a* and lower bite guide **212***b* are spaced apart such that a subject biting down on the upper bite guide **212***a* with the subject's upper jaw and biting up on the lower bite guide **212***b* with the subject's lower jaw opens the subject's oral cavity to inspection through the opening **211**.

[0052] In the illustrated embodiment, the mouthpiece is curved out of the plane of the view of FIG. **2**A to follow a typical shape of a face of a subject so that left side, right side and middle of the opening **211** in the mouthpiece are about equally far from the center of the surface of an oral cavity of a subject, where a back of the tongue of the subject normally leads to the top of the throat of the subject. In some embodiments, the mouthpiece is flat. Curving the mouthpiece offers the advantage of changing the angle of view of the imaging, so that regions of the buccal mucosa (inner part of cheek) that are near the lips can be seen. A scanner without a curved surface would miss this region.

[0053] In some embodiments, the posterior side of the mouthpiece **210** includes one or more sources of light in each of one or more light banks. In the illustrated embodiment, the mouthpiece includes four light banks **216***a*, **216***b*, **216***c* and **216***d*, collectively called light banks **216** hereinafter. In some embodiments, the light banks (e.g., light banks **216**) include bright light sources with a wide wavelength band encompassing most or all of the visible spectrum. In some embodiments, one or more of the light banks **216** include a light source for exciting fluorescence emission from the tissue of the oral cavity of the subject, such as a blue LED at 488 nm wavelength to excite auto-fluorescence in human tissue of the oral cavity. Such auto-fluorescence indicates variations of health of the tissue within the oral cavity. In some embodiments, the mouthpiece includes a power source for the light banks **216** such as a battery pack. In the illustrated embodiment, the mouthpiece includes a power cord **218** that extends from the light banks **216** of the mouthpiece **210** to a power source, such as a battery pack, that resides off the mouthpiece **210**.

[0054] In various embodiments, the mouthpiece **210** includes a space between the opening and the bite guides **212** and light banks **216** to allow a flange of a camera mount, described below, to slide along the opening. In some embodiments, the portion of the mouthpiece on the upper and lower edge of the opening, along which the flange slides, includes an upper rail **214***a* and lower rail **214***b*, respectively (collectively referenced hereinafter as rails **214**) configured especially for this purpose. For example, in some embodiments, the rails **214** include a reinforced material, such as metal, or especially smooth coating, such as TEFLON™ produced by DUPONT CO.™ of Wilmington, Del., or other material, or some lubricant as may be well known in the art, or some combination. In embodiments using either nylon or ABS as the material of the mouthpiece **210**, canola oil is advantageous as a lubricant because it is cheap and non-toxic. Some embodiments used thin metal as well as magnets (magnets were used in some embodiments to provide a continuously loaded bearing for smooth scanning, which especially helps with pushbroom imaging).

[0055] In some embodiments, a portion of the mouthpiece, such as a portion along the left or right side of the opening **211**, is hinged to allow the flange **222** of the camera mount **220** to be inserted into the opening. In some embodiments, the rail is kept going throughout the device so that the flange just inserts into one of the sides.

[0056] The bracket **200** also includes a camera mount **220** that is configured to hold, on the opposite side (the anterior side facing the clinician) a camera to capture an image through the opening **211** in the mouthpiece **210**. The posterior side of the camera mount **220**, depicted in FIG. **2**A, includes a flange **222** to engage the mouthpiece **210** along the opening **211**, e.g., along rails **214**, in such a way that the flange **222**, and camera mount **220** of which the flange is part, can slide along the opening **211** of mouthpiece **210**, as indicated by the dashed arrow in FIG. **2**A. In some embodiments, the flange **222** is shaped so that when the camera mount **220** is rotated, as indicated by the curved arrow, the flange **222** disengages from the opening **211** of the mouthpiece **210**; and, the camera mount **220** can be removed. The camera mount **220** can be engaged with the mouthpiece **210** by reversing this action. For example, when engaged, the flange **222** is larger in the vertical direction than the horizontal direction; e.g. the flange **222** is shaped as an oval with major axis aligned along the vertical when engaged with the opening **211**, or shaped as a rectangle with the longer dimension aligned with the vertical when engaged with the opening **211**.

[0057] The posterior side of the camera mount **220** includes an optical path **223** through to the anterior side of the camera

mount at a position within the opening **211** so that light can pass from the oral cavity through the opening **211** and optical path **223** to the anterior side of the camera mount **220**. Thus the optical path is configured for light to pass through the camera mount and through the opening in the mouthpiece. In some embodiments, the optical path is disposed beside the flange **222**. In the illustrated embodiment, the optical path **223** passes through the flange **222**.

[0058] In some embodiments, the optical path includes a pin hole that provides wide angle focus for all surfaces of the subject's oral cavity regardless of the varying distances from the pin hole to the surface of the subject's oral cavity. In some embodiments, a material lens, e.g., made of glass or plastic, is included in the optical path in addition to or instead of the pin hole. For example, in some embodiments, a fish eye lens, which expands the center of the field of view and compresses the edges, is included in the optical path.

[0059] In some embodiments, the optical path **223** includes a filter **225** that blocks light in a wavelength band corresponding to the fluorescence excitation light and passes light in a fluorescence emission wavelength band. In some embodiments, the filter **225** is configured to be moved into and out of the optical path **223**. In such embodiments, the bracket **200** may be used to image both a bright light field (bright field) and a fluorescence emission field.

[0060] In some embodiments, the power source for the light banks **216** on the mouthpiece **210** is disposed on the camera mount **220**. For example, as depicted, the power source is a battery pack **228** disposed on a posterior side of the camera mount **220**; and, the power cord **218** for the light banks **216** on the mouthpiece **210** is connected to the battery pack **228**.

[0061] FIG. **2**B depicts the anterior view of the oral cavity image bracket **200**. The mouthpiece **210**, opening **211**, rail **214***a*, rail **214***b*, camera mount **220** and optical path **223** are as described above for FIG. **2**A.

[0062] The anterior side of the camera mount **220** includes one or more clips, such as adjustable clip **226***a* **226***b* and **226***c* (collectively referenced hereinafter as adjustable clips **226**), for removably securing a camera to the anterior side of the camera mount **220**. The adjustable clips **226** are disposed in tracks **227***a*, **227***b*, **227***c*, respectively, (collectively referenced hereinafter as tracks **227**) so that the adjustable clips **226** can be moved along the tracks **227** to acquire a position useful for holding a camera in place so that the camera is positioned to capture an image through the optical path **223**. In other embodiments, one or more clips are configured differently. For example, in some embodiments, the camera mount **220** is configured for a certain shaped camera, such as a camera cell phone, and the clips are not adjustable along a track, such as tracks **227**. In some embodiments, each of one or more clips comprises a ledge extending out of the plane of FIG. **2**B and a small plate perpendicular to the ledge at the outer edge of the ledge. Thus, the one or more clips are configured to removably hold a camera on the anterior side of the camera mount to record light passing through the optical path from the posterior side of the camera mount. In various embodiments, the clip is configured to removably hold a camera selected from a group comprising: a film camera; a digital camera; a digital camera with on board processor; a cell phone with digital camera; a programmable cell phone with digital camera, among others.

[0063] FIG. **2**C depicts an example oral cavity image system **250**, according to an embodiment. The illustrated system **250** includes the oral cavity image bracket **200**, a cell phone **280** with digital camera, a communications network **290**, and an oral cavity image server **292**. In other embodiments, a digital camera is used instead of cell phone **280**, or the communications network **290** and remote server **292** are omitted, or the system is changed in some combination of ways.

[0064] The oral cavity image bracket **200** is depicted in perspective and includes mouthpiece **210** and camera mount **220**. The flange **222**, optical path **223**, adjustable clip **226***b*, power cord **218** and battery pack **228** are as described above. In the illustrated embodiment, the battery pack **228** includes a switch **229** configured to turn power on and off to the light banks **216** on mouthpiece **210**.

[0065] The cell phone **280** includes a digital camera as is common in modern smart phones, such as mobile terminal **1301** described in more detail below with reference to FIG. **13**. Light passing into the cell phone camera aperture **284** is captured by a sensor plane, such as a CCD array **1365** depicted in FIG. **13**, and stored in the memory, e.g., memory **1351** depicted in FIG. **13**, of the cell phone **280**, e.g., mobile terminal **1300** of FIG. **13**. The cell phone **280** includes an oral cavity image application **230** configured to execute on a microprocessor main control unit (MCU) of the cell phone **280**, such as the MCU **1303** depicted in FIG. **13**. The instructions for the oral cavity image application **230** are stored in memory **1351** until used by the MCU **1303**.

[0066] The cell phone **280** also includes a transceiver, such as transceiver **1315** depicted in FIG. **13**, for radio frequency communications with a communications network **290**, such as a cell phone network, through wireless connection **291**. In some embodiments, the oral cavity image application **230** is downloaded from an oral cavity image server **292** through the communications network **292** to the cell phone **280**. In some embodiments, the oral cavity image application **230** performs all the processing to determine the standard image, for both bright field and fluorescence images, based on the light captured through the cell phone aperture **284**. Both the captured images and the standardized image are displayed on the cell phone **280**, e.g. on display **1307** depicted in FIG. **13**. In some embodiments, some or all of the processing to determine the standard image based on the light captured through the cell phone aperture is performed by the remote oral cavity image server **292**; and, image data is transmitted from the cell phone **280** to the server **292** through the communications network **290**, and the resulting standardized image, whether a bright field or fluorescence image, is returned by the oral cavity image server **292** through the communications network **292** to the cell phone **280**, where the standard image or images are displayed.

[0067] During operation of the system **250**, the camera, such as cell phone **280**, is a loaded onto the camera mount **220** as indicated by the dashed arrow in FIG. **2**C. In some embodiments the adjustable clip **226***b* is replaced by a fixed shaped piece of rigid material, against which the cell phone may lodge.

[0068] Although processes, equipment, and data structures are depicted in FIG. **2**C and following drawings as integral blocks in a particular arrangement for purposes of illustration, in other embodiments one or more processes or data structures, or portions thereof, are arranged in a different manner, on the same or different hosts, in one or more databases, or are omitted, or one or more different processes or data structures are included on the same or different hosts. For example, in some embodiments, a processor (not shown) is included on the oral cavity image bracket **200**, such as on camera mount

220; and, the oral cavity image application **230** executes, in whole or in part, on the processor included on the oral cavity image bracket **200**. In such embodiments, a wired (e.g., a universal serial bus, USB, cable) or wireless (e.g., Bluetooth) connection is established between the camera removably attached to the camera mount **220** and the processor on the bracket **200**.

[0069] FIG. **3**A, FIG. **3**B and FIG. **3**C are block diagrams that illustrate example fields of view from multiple positions of a camera mount **310** with cell phone attached along an opening in a mouthpiece **210** of the example oral cavity image bracket **200**, according to an embodiment. Each of FIG. **3**A, FIG. **3**B and FIG. **3**C shows operation of the bracket **200** from overhead looking down on the subject. A head **390** of the subject, such as a patient, is indicated with a mouthpiece **210** in place as the subject bites down on the bite guides **212**. The oral cavity **392** of the subject is indicated by the dashed curve opening to the mouthpiece **210**. The camera mount **310** with cell phone clipped in place is also depicted.

[0070] FIG. **3**A depicts the camera mount **310** with cell phone in a first position farthest to the subject's left but still within the opening of the mouthpiece **210**. The field of view **320***a* of the cell phone camera is depicted as a dotted triangle that captures light from the subject's right side and center of the subject's oral cavity **392**. FIG. **3**B depicts the camera mount **310** with cell phone in a second position within the middle of the opening of the mouthpiece **210**. The field of view **320***b* of the cell phone camera is depicted as a dotted triangle that captures light from the center of the subject's oral cavity **392**. FIG. **3**C depicts the camera mount **310** with cell phone in a third position farthest to the subject's right but still within the opening of the mouthpiece **210**. The field of view **320***c* of the cell phone camera is depicted as a dotted triangle that captures light from the subject's left side and center of the subject's oral cavity **392**.

[0071] FIG. **3**D is a block diagram that illustrates example overlapping raw image frames captured from the multiple positions of the camera mount, according to an embodiment. The projection of the target walls of the oral cavity **394** on the plane of the camera sensor is indicated by the dashed lines. The portion of that image captured by the cell phone camera in one position, e.g. the position of FIG. **3**A, is shown as raw frame **330***a*. The portion of that image captured by the cell phone camera in another position, e.g. the position of FIG. **3**B, is shown as raw frame **330***b*. The portion of that image captured by the cell phone camera in yet another position, e.g. the position of FIG. **3**C, is shown as raw frame **330***c*. It is these multiple raw frames, such as frames **330***a*, **330***b* and **330***c* (collectively referenced hereinafter as raw frames **330**), which are merged together to generate a standardized image by the oral cavity imaging processing method, such as oral cavity image application **230**, and, optionally, oral cavity image server **292**.

[0072] In some embodiments, the camera mount **220** is rotated several degrees in a plane perpendicular to the optical path while the flange **222** is still engaged with the mouthpiece **210** through the opening **211**. In such embodiments, the raw frames include frames with a rotated view of the target walls of the oral cavity **394**, compared to the views depicted in FIG. **3**D; and, these raw frames are also included in the merge processing by the oral cavity image application **230** or oral cavity image server **192** or both. This can allow getting a wider field of view which can help get peripheral features (landscape mode). In portrait mode, it enables a broader amount of the hard palate and floor of the mouth (area under tongue).

[0073] Thus, in various embodiments, pioneering plug-and-play geometrical optics are deployed for use with camera-phones for medical applications; specifically targeted towards both bright field and fluorescence imaging of the oral cavity. The oral cavity presents a complex three dimensional surface making it extremely difficult to take images with comprehensive coverage that could be used as a standard for imaging the cavity. In fact, no prior standard (or platform) exists that can image the entire oral cavity. Embodiments of the present invention are anticipated to find great utility in field settings with shortage of doctors and oral surgeons, where clinicians often rely on memory to assess the extent of oral hygiene problems in patients of rural areas in many countries, e.g., because of excessive tobacco consumption.

2. Method

[0074] FIG. **4**A is a flow chart that illustrates an example method **400** for producing and using one or more standard images of the oral cavity utilizing the oral cavity image bracket and a cell phone with built in camera and processor, according to an embodiment. Although steps are depicted in FIG. **4**A, and in subsequent flowchart FIG. **4**B, as integral steps in a particular order for purposes of illustration, in other embodiments, one or more steps, or portions thereof, are performed in a different order, or overlapping in time, in series or in parallel, or are omitted, or one or more additional steps are added, or the method is changed in some combination of ways.

[0075] In step **401**, an oral cavity image application, such as oral cavity image application **230**, is loaded into a cell phone equipped with digital camera, also known as a camera phone. Any method may be used to load the application into the camera phone, such as downloading the application from a remote server, such as the oral cavity image server **292**. In some embodiments, the oral cavity image application **230** is loaded into the cell phone **280** through a cable connection to a computer, such as depicted in FIG. **11**. In some embodiments, the oral cavity image application **230** is loaded into the cell phone **280** from a removable memory card, such as a SIM card **1349** depicted in FIG. **13**. In some embodiments using a processor on the bracket **200** instead of the processor on the camera, step **401** includes loading the oral cavity image application **230** onto that processor on the bracket **200**.

[0076] In step **403**, the cell phone equipped with a camera and processor is removably attached to the camera mount of the oral cavity image bracket, e.g. using one or more clips on the camera mount such as one or more adjustable clips **236**. The clips are configured to hold a camera, such as the cell phone **280**, on the anterior side of the camera mount to record light passing through the optical path from the posterior side of the camera mount. In other embodiments, a different camera is mounted to the camera mount **220** in order to capture light transmitted through the optical path **223**.

[0077] In step **405** the oral cavity image application **230** is executed, which prompts a user to enter patient and image metadata information into the application. In the illustrated embodiment, the application **230** executes on the cell phone **280** used as camera; and, the user interface of the cell phone **280** is used to prompt for and receive the metadata information. In other embodiments, a separate processor on the

bracket **200** prompts for or receives the metadata or both through a separate interface, or using commands exchanged with the camera.

[0078] In step **407** the next subject (e.g., patient) is caused to bite down on the bite guides of the mouthpiece (e.g., on bite guides **214** of mouthpiece **210**) to expose the subject's oral cavity to inspection through the opening (e.g., opening **211**) in the mouthpiece (e.g., mouthpiece **210**). In some embodiments, step **407** includes sanitizing the mouthpiece after use by one patient, or after prolonged non-use, for reuse with the next patient.

[0079] In step **409**, it is determined whether fluorescence imaging is to be performed. In some embodiments this determination is made automatically. For example, a bright field image is automatically followed by a fluorescence image; or, every image is a bright field image; or, every image is a fluorescence image. In some embodiments, this determination is made by an operator.

[0080] If it is determined in step **409** that fluorescence imaging is not performed, then in step **411** the filter is positioned out of the optical path, either automatically or by manual operation by the operator/clinician. In some embodiments the filter, e.g., filter **225**, is configured to pass only fluorescence emissions excited by light of wavelength near 488 nm. In step **412**, the light banks of bright light are turned on. In some embodiments, as described in more detail below, the light banks **416** of bracket **200** are omitted, and the light of the camera, such as light source **1361** depicted in FIG. **13**, of the cell phone is used as the bright field light source. Control then passes to step **417** described below.

[0081] If it is determined, in step **409**, that a fluorescence image is to be performed, then in step **413** the filter is positioned in the optical path, again either automatically or by manual operation by the operator/clinician. In step **414**, the fluorescence excitation light sources of the light banks **416** of bracket **200**, e.g., blue LED lights, are turned on. In some embodiments, light banks **416** of bracket **200** are omitted; and, the light of the camera, such as light source **1361**, of the cell phone is used as the light source. In some such embodiments, a second filter is placed over the light source of the cell phone to pass only the fluorescence excitation wavelengths, e.g., at 488 nm for auto-fluorescence. Control then passes to step **417**.

[0082] In step **417**, the camera mount **220** is slid along the opening **211** of the mouthpiece **210** to the first or next position for capturing a raw frame image of the subject's oral cavity. The camera, such as cell phone **280**, is then operated to capture the raw frame image. The raw frame image is stored in the memory of the camera, such as the cell phone **280** (e.g., in memory **1351**), or in a separate processor that is part of the bracket **200**, or some combination.

[0083] In step **419**, it is determined whether a raw image at another position is desired. If so, control passes back to step **417** to slide the camera mount to the next position along the opening of the mouthpiece and capture the next raw frame image using the camera, such as the cell phone **280**. If no further raw images are desired in the current imaging mode (either bright field or fluorescence imaging), then control passes to step **421**.

[0084] In step **421**, it is determined whether another mode of imaging is desired. For example, it is determined whether a set of raw frame images in a bright field mode have been collected, but raw frame images in a fluorescence mode are still desired. If so, control passes back to step **409**, described above. If not, control passes to step **423**.

[0085] In step **423**, the light banks are turned off and the oral cavity image application **230** is executed to merge several raw frame images into a standard image of the oral cavity of the subject. One or more standard images are produced, e.g., a bright field standard image and an auto-fluorescence standard image. In some embodiments, the standardized image is a set of standardized views with the subject's tongue in different positions. The process performed by the oral cavity image application **230** is described in more detail below with reference to FIG. **4**B.

[0086] In step **425**, the one or more standard images of the oral cavity of the subject are displayed for and analyzed by the operator/clinician. In addition, the raw frames or standard images or metadata or some combination is stored, either locally on the memory of the camera, such as memory **1351** of cell phone **280**, or remotely on the oral cavity image server **192**, or some combination. In some embodiments, step **425** includes recommendations by an automated algorithm on which areas of the image are suspect and worth a close examination by the clinician, or an estimated volume of a diseased areas, such as a lesion.

[0087] In step **427**, it is determined whether there is another subject, such as another patient, to be examined using the oral cavity image system, such as system **250**. If so, control passes back to step **405** described above. If not, the process **400** ends.

[0088] FIG. **4**B is a flow chart that illustrates an example computer program method **440** for producing a standard image of the oral cavity based on overlapping image frames captured from the multiple positions of the camera mount, according to an embodiment. The method **440**, in various embodiments, is performed by oral cavity image application **230** on the digital camera or camera phone or on a separate processor on the bracket **200**, or by the remote oral cavity image server **292**, or some combination. The method **440** uses two characteristic image manipulations: robust image mosaic formation; and, depth map based three dimensional (3-D) mesh generation for an individual oral cavity. Either one of the methods can be applied independently or in combination. The two methods can be applied for either a single image or multiple series of images.

[0089] In step **441**, a prompt is sent to a display for the user to enter patient and image metadata, such as patient name, date and time, location, among others. Several screens of a graphical user interface (GUI) capable of such prompting are illustrated below with reference to FIG. **6**A through FIG. **6**C. In some embodiments, step **441** includes prompting the user and receiving input that indicates whether a pinhole or material lens is being used in the optical path. The former provides an in focus image for all portions of the surface of the oral cavity, while the latter focuses at a depth of field that can be used to infer the three-dimensional contours of the surface of the oral cavity, as described in more detail below. In some embodiments, step **441** includes prompting the user and receiving input that indicates whether fixed or dynamic lighting is to be used. In general, fixed lighting is attached to the mouthpiece, as depicted above as the light banks **216** in FIG. **2**A; while, dynamic lighting is attached to the camera mount, e.g. camera mount **220** depicted in FIG. **2**A. With dynamic lighting, the light intensity of a portion of the surface of the oral cavity is different for different frames captured by the camera on the camera mount.

[0090] In step **443**, the next image frame is captured and received for the current patient and current mode. As used here "mode" refers to different lighting and optical characteristics of the capturing process, such as capturing a bright field image or an auto-fluorescence image. In some embodiments, step **443** includes detecting user selection of an active area on the GUI, such as a button labeled "take picture," that corresponds to capturing an image. The color and intensity measurements received at the optical sensor array are stored as image frame data, such as in a memory device in the camera, camera phone, or separate processor. In embodiments using a separate processor, image frame data captured at the camera is transmitted to the separate processor, either wirelessly or through a connected cable.

[0091] In step **445**, it is determined whether the current image frame was taken with a material lens, such as through an optical path that includes a removable fish eye lens. If not, control passes to step **451**, described below. If so, then control passes to step **447** to determine the distance of each pixel from the digital camera based on the focus. Any method known in the art to infer distance from focus and properties of the focusing lens may be used. In one embodiment, a depth map is constructed during step **447** using either single images or mosaic images utilizing algorithms described by Ashutosh Saxena, Sung H. Chung, Andrew Y. Ng, "Learning Depth from Single Monocular Images," in *Neural Information Processing Systems* (NIPS) v18, 2005, and Ashutosh Saxena, Min Sun, Andrew Y. Ng, "Make3D: Learning 3D Scene Structure from a Single Still Image," *IEEE Transactions of Pattern Analysis and Machine Intelligence* (PAMI), vol. 30, no. 5, pp 824-840, 2009, the entire contents of each of which are hereby incorporated by reference as if fully set forth herein, except for terminology that is incompatible with that used herein. In this embodiment, pixels that are in focus show a different net intensity than pixels out of focus. Given the characteristics of the lens, the distance to the pixels in focus can be determined.

[0092] As a result of this computation, the depths of the pixels in focus can be added to a 3-D model of the surface of the oral cavity and combined with other image frames, as described below, to produce a full 3-D mesh surface for the surface of the oral cavity. The advantage of applying this image processing method to oral cavity images comes from the fact that real geometrical quantities can be measured in metric units. Thus it is easy to quantify if a detected lesion is growing or shrinking. Consequently, automated marking and clustering of lesions based on size is possible. Control then passes to step **451**.

[0093] In step **451**, it is determined whether dynamic lighting is used during image frame capture, for example based on response to prompts in step **441**. If not, control passes to step **461**, described below. If so, control passes to step **453** to correct pixel intensity values on the frame based on the imaging geometry, such as the angle of illumination and the depth, if known, of the pixel. Control then passes to step **461**.

[0094] In step **461**, it is determined whether there is one or more previous image frames. In some embodiments, step **461** through step **473** are performed only after all frames have been captured for the current patient. In the illustrated embodiment, step **461** through step **473** are performed after each frame is captured. If there are no previous image frames, then control passes to step **471**, described below. If there are one or more previous image frames, then control passes to step **463**.

[0095] In step **463**, the current image frame is registered with respect to the previous image frame using any registration method available known in the art. The step provides information that associates one or more pixels in a previous image frame with corresponding one or more pixels in the current image frame. In step **467**, values for each pixel are determined based on the values in all corresponding pixels from any previous image frames. In one embodiment, the pixels are analyzed by column of pixels. For each pixel in the column, in this embodiment, the value is selected from a single pixel determined to be in focus, either because it was captured using a pinhole or was determined to be in focus during step **447**, described above. In other embodiments, other procedures are used to combine the information from the corresponding pixels. For example, in some embodiments a median pixel value is selected; while, in other embodiments, a weighted or unweighted average of the corresponding pixels from two or more image frames is used as the pixel value. Control then passes to step **471**. In an illustrated embodiment, the image mosaic is performed by algorithms using invariant features. See, for example, M. Brown and D. Lowe, "Automatic panoramic image stitching using invariant features," *International Journal of Computer Vision*, v74 no. 1, pp 59-73, 2007, the entire contents of which are hereby incorporated by reference as if fully set forth herein, except for terminology inconsistent with that used herein.

[0096] In step **471**, it is determined whether the current image frame was captured in the fluorescence mode. If not, control passes to step **475**, described below. If so, control passes to step **473**. In step **473**, pixels are classified as either normal or abnormal based on the relative intensity of the pixel values. Abnormal or diseased tissues fluoresce at a lower intensity than healthy tissue. Control then passes to step **475**.

[0097] In step **475**, it is determined whether there is another frame to be captured from the same patient. If so, control passes back to step **443**, described above. If not, control passes to step **481**. For example, the user selects an active area of the GUI to indicate completion of scanning the current patient, then control passes to step **481**.

[0098] In Step **481**, one or more standard images of the oral cavity in bright field or fluorescence mode, or both, are generated and displayed on the GUI, such as on the display of the cell phone. If a 3-D mesh has been computed, in some embodiments the 3-D mesh is optionally displayed over the image.

[0099] In step **483**, the state of the tissue displayed in the one or more standard images is determined. For example based on the relative values within the image, or between the bright field and fluorescence images, or user input, or some combination, the healthy tissue is distinguished from the diseased tissue. In some embodiments that use the 3-D mesh, the volume of diseased tissue, such as the volume of a tumor, is determined during step **483**.

[0100] In step **485**, the one or more standard images and metadata, including any automated analysis results, are stored in memory, either on the local device or on the remote server **292**. In step **487**, it is determined whether there is another patient whose oral cavity is to be scanned. If so, control passes back to step **411**. If not, then the process ends.

[0101] Thus, the method **440** uniquely combines mosaic based imaging (using one or more algorithms known to one of ordinary skill) and depth perception and 3D mesh generation from the same frames (using one or more different algorithms known to one of ordinary skill) for the special circumstances

of oral screening. Collecting specified angular frames of known geometry and preparing the data for combining these two methods together allow a universal metric coordinate system that can be used to mark, measure and track lesions. Identification of lesions is done based on various threshold algorithms (using one or more still different algorithms known to one of ordinary skill) on this uniquely prepared data set.

3. Example Embodiments

[0102] FIG. **5**A and FIG. **5**B are photographs that illustrate an example oral cavity image bracket and system, according to another embodiment. FIG. **5**A is a photograph that depicts the posterior view of oral cavity image bracket **500** according to this embodiment. This embodiment includes mouthpiece **510** and camera mount **520**. The mouthpiece **510** includes upper bite guide **512** that is made up of both a groove and a ledge with a plate turned up at the outer edge of the ledge. A similar lower bite plate is also shown. Between the upper and lower bite guides is mouthpiece opening **511**. The mouthpiece **510** also includes light banks **516** that are illuminated and power cord **518** that connects to battery pack **528** on camera mount **520**. The camera mount **520** includes flange **522** and optical path **523** as well as battery pack **528**. Visible through an opening in the camera mount **520** is the dark body of cell phone **580** used to capture light that passes through optical path **523**.

[0103] FIG. **5**B is a photograph that depicts the oral cavity image bracket **500** in a lateral view during operation. Apparent in FIG. **5**B is the mouthpiece **510** on which subject **590** is biting so as to present a view of the subject's oral cavity through opening **511**. Also visible is the camera mount **520** engaged with the opening **511** of mouthpiece **510** by flange **522**. Also visible is cell phone **580** removably attached to the camera mount **520**, so that light passing through the optical path of camera mount **520** is captured by the camera aperture of cell phone **580**. In some embodiments, portions of the camera mount **520** serves as a phone case. This type is manufactured differently for each type of phone. In some embodiments, the camera mount includes a flat substrate (such as a circuit board with a hole in it and lighting and battery attached, as described below with reference to FIG. **10**D) that attaches via double sided office tape or double sided foam.

[0104] FIG. **5**C and FIG. **5**D are photographs that illustrate an example oral cavity image bracket **550** and system, according to yet another embodiment. FIG. **5**C is a photograph that depicts a lateral view of oral cavity image bracket **550** according to this embodiment. This embodiment includes mouthpiece **560** and camera mount **570**. The mouthpiece **560** includes mouthpiece opening **561**. The mouthpiece **560** does not include light banks or a power cord or a battery pack. The camera mount **570** includes flange **572** and an opening **571** in the mount **570**, which serves both as an optical path and a path for the light from a light source (e.g., light source **1361**) of the camera (e.g., mobile terminal **1300**) to illuminate the oral cavity. Also visible is the dark body of cell phone **584** used to capture light that passes through the opening **571** in the camera mount **570**. This opening **571** in the camera mount **570** permits use of the cell phone camera aperture and cell phone light source in lieu of an optical path through the mount and light banks on the mouthpiece. In the illustrated embodiment, the flange **572** comprises two separate portions, an upper portion and a separate lower portion, and the opening **571** is disposed between the upper and lower portions of flange **572**.

[0105] It is important to note the difference between the electronics setups in FIG. **5**C and FIG. **5**A. Some embodiments (e.g., FIG. **2**A and FIG. **5**A) utilize static lighting, wherein the illumination in the oral cavity does not change as the scan progresses. In the other embodiment (FIG. **5**C) dynamic lighting occurs (the lighting moves as the scan progresses). Both modes have advantages and disadvantages. In static lighting, shadows stay put (which could make an area look suspicious in fluorescence). In dynamic lighting, it is possible to tell shadows from real suspicious regions; however, the intensity of lighting in a region changes in each image which can be problematic when tracking features

[0106] FIG. **5**D is a photograph that depicts the oral cavity image bracket **550** as an anterior view during operation. Apparent in FIG. **5**D is the mouthpiece **560** on which subject **592** is biting so as to present a view of the subject's oral cavity through opening **561**. Also visible is the camera mount **570** engaged with the opening **561** of mouthpiece **560** by flange **572**. Also visible is cell phone **584** removably attached to the camera mount **570**, so that light **583** from the cell phone light source illuminates the oral cavity of subject **592** and light passing through the optical path of camera mount **570** is captured by the camera aperture of cell phone **584**.

[0107] FIG. **6**A through FIG. **7** are diagrams of user interfaces utilized in the processes of FIG. **4**A, according to various embodiments. These figures illustrate an example graphical user interface for an oral cavity image application for a programmable cell phone with built-in camera and processor, according to another embodiment. The screen includes one or more active areas that allow a user to input data or operate the application. As is well known, an active area is a portion of a display to which a user can point using a pointing device (such as a cursor and cursor movement device, or a touch screen) to cause an action to be initiated by the device that includes the display. Well known forms of active areas are stand alone buttons, radio buttons, check lists, pull down menus, scrolling lists, and text boxes, among others. Although areas, active areas, windows and tool bars are depicted in FIG. **6**A through FIG. **7** as integral blocks in a particular arrangement on particular screens for purposes of illustration, in other embodiments, one or more screens, windows or active areas, or portions thereof, are arranged in a different order, are of different types, or one or more are omitted, or additional areas are included or the user interfaces are changed in some combination of ways.

[0108] In FIG. **6**A, the graphical user interface (GUI) of a display (e.g. **1307**) of the cell phone (e.g. mobile terminal **1300**) includes a first page **601** with a control panel **691** for operating the cell phone, and, specific to the oral cavity image application, e.g., application **230**, a label **609** indicating the oral cavity scan application (Oscan in the illustrated embodiment). The page **601** also includes labels **610***a* through **610***h*, text boxes **612***a* through **612***e*, and pull down menus **614***a* and **164***b*. For example, label **610***a* includes the text "Please Enter Name (NO SPACES):" and text box **612***a* is configured to receive text entered by a user/clinician that indicates a name for the subject/patient. Similarly, label **610***b* includes the text "Device Model (NO SPACES):" and text box **612***b* is configured to receive text entered by a user/clinician that indicates the model of camera, digital camera or camera phone used to scan the oral cavity. Similarly, label **610***c* prompts for "Age" of the patient and text box **512***c* is configured to receive that information. Labels **610***d* and **610***e* prompt for "Gender" and "Race" of the patient and pull down menus **614***a* and **614** are

configured to present a list of choices approved to receiving the data. Labels **610***f* and **610***g* prompt for "Tobacco history (Pack Years)" and "Alcohol History," while text boxes **612***d* and **612***e*, respectively, are configured to receive that data. A user can scroll down page **601** using normal controls for the device, such as sliding a finger along a touch screen display on the device. Text boxes off screen are scrolled to in order to receive other information that the user/clinician chooses to enter, such as the "Additional Patient information" prompted by label **610***h*.

[0109] The information provided by the user/clinician in the text boxes and pull down menus of page **601** constitutes metadata for the raw images to be captured and the standardized image to be generated. Patient background data is often just as important as the images collected. On other pages of the GUI, the user/clinician is prompted for additional information which, upon entry by the user/clinician, becomes metadata about the raw images captured and the standardized image produced. In some embodiments, an open source software tool (e.g., ANDROID™ based software tools from GOOGLE INC™ of Mountain View, Calif.) is used to allow for integrated GUI, metadata collection, and data management for a cell-phone based oral scanner.

[0110] In FIG. **6**B, the graphical user interface (GUI) of a display (e.g. **1307**) of the cell phone (e.g. mobile terminal **1300**) includes another page **602** used for operating the OScan application, a particular embodiment of the oral cavity image application **230**. Besides control panel **691** and application label **609**, described above, page **602** includes pull down menu **624** and buttons **626***a* through **626***c*. The pull-down menu **624** lets the user/clinician indicate which frame is being captured among left, center, right, among others as listed in FIG. **6**C, described below. Button **626***a*, labeled "Press to take photos," is configured to be activated by the user/clinician to cause the camera (e.g., mobile terminal **1300**) to capture an image frame. Button **626***b*, labeled "Retake Photo," is configured to be activated by the user/clinician to cause the camera (e.g., mobile terminal **1300**) to delete the last image frame and capture a new image frame. Button **626***c*, labeled "Next person," is configured to be activated by the user/clinician to cause the camera (e.g., mobile terminal **1300**) to finish storing image frames and metadata for one subject so that such information can begin to be collected for another subject (e.g., to follow the YES branch from step **487** described in FIG. **4**B, above.

[0111] In FIG. **6**C, the graphical user interface (GUI) of a display (e.g. **1307**) of the cell phone (e.g. mobile terminal **1300**) includes yet another page **603** used for operating the OScan application. Besides control panel **691** and application label **609**, described above, page **603** includes labels **630***a* through **630***g* and corresponding radio or toggle buttons **638***a* through **638***g*. The labels **630***a* through **630***g* indicate multiple locations for frames to capture, which are recommended for producing the standardized image. In the illustrated embodiment, seven locations indicated by the labels **630***a* through **630***g* are "Left (L)," "Left of Center (LC)," "Center (C)," "Right of Center (RC)," "Right (R)," "Tongue Left (TL)," "Tongue Right (TR)," respectively. The toggle buttons are filled, as depicted for toggle button **638***a*, as each frame at the corresponding location is captured and stored by the application on the camera, digital camera or camera phone, other equivalent device.

[0112] FIG. **7** is a block diagram that illustrates an example browser graphical user interface **700** for a remote server (e.g., remote oral cavity image server **292**) configured for processing oral cavity images, according to another embodiment. As is well known in the art, a browser is a client process operating on a local device that interacts over a network with a server process using a particular communication protocol called the hypertext transfer protocol (HTTP) and a particular language to indicate formatting called the hypertext markup language (HTML). In the illustrated embodiment, browser interface **700** displays a captured raw image **720** (e.g., on display device **1114** of computer **1100** depicted in FIG. **11**). The illustrated browser interface **700** also displays multiple thumbnails of frames captured **710** which are active areas the user/clinician can select to choose the raw image to display as captured raw image **720**. Other GUI control active areas **720** are also presented in the browser interface **700** to allow the user/clinician to invoke various functions provided by the remote oral cavity image server **292**. For example, in some embodiments, the GUI control active areas **720** include an active area to cause the remote oral cavity image server **292** to merge two or more of the raw images selected from the thumbnails of frames captured **710**. In the illustrated embodiment, the captured raw frame image **720** and thumbnails of raw frames captured **710** are all bright field images in full color. In other embodiments, one or more captured raw frame image **720** and thumbnails of frames captured **710** are fluorescence raw frame images.

[0113] FIG. **8** is an image that illustrates an example standardized fluorescence image **800** for an oral cavity, according to an embodiment. The image **800** is a composite of multiple raw frame fluorescence images merged during step **423** of method **400**, e.g. according to the method **440** of FIG. **4**B. The image **800** is an image mosaic for comprehensive imaging of the oral cavity. The entire scan takes less than 1 minute to perform. As demonstrated by the above embodiment, a low-cost scanner for auto fluorescence imaging using a NEXSUS™ Phone from GOOGLE INC.™ as been implemented.

[0114] FIG. **9**A through FIG. **9**F are images that illustrate example bright field images for an oral cavity, according to an embodiment. These images are raw bright field images collected using a cellphone camera and its embedded optics only, without added fisheye lens or external pinhole. These are the standard positions that an oral specialist will place the tongue during an oral examination. These images are not merged, but the mosaic shows the high risk areas for cancer; and presents all of high risk areas for oral cancer in a single mosaic. Each standardized image may comprise one or more of these views after merging several raw frames.

[0115] FIG. **10**A through FIG. **10**F are scaled drawings for a mouthpiece and camera mount, according to a particular embodiment. FIG. **10**A depicts in oblique view a camera mount housing **1020** for a camera mount, such as camera mount **220**. The housing **1020** includes a flange **1022** and optical path opening **1023**. FIG. **10**B depicts in a corresponding oblique view a mouthpiece **1010**, such as **210**. The mouthpiece **1010** includes opening **1011** and upper and lower bite guides **1012**. FIG. **10**F depicts mouthpiece **1010** in an opposite oblique view.

[0116] FIG. **10**C depicts manufacturing drawings for the housing **1020**, including aligned top view **1020***a*, front view **1020***b*, and side view **1020***c*, for a particular embodiment. In this embodiment, the distance associated with various dimensions are given in millimeters (mm, 1 mm=10^{-3} meters). Thus, the housing width is about 37 mm, height is about 69 mm, opening **1023** is about 32 mm wide and 26.15 mm high.

Each flange **1022** is about 15 mm wide, 2.23 mm high, and extends perpendicular to the face of the housing **1020** by up to 3.36 mm and curved with a radius of curvature of about 33 mm. the opening **1023** is set back from the face of the housing by about 2.5 mm, and the housing is about 8 mm thick. An opening in the side of the housing is about 8.75 mm from front to back and about 7 mm high.

[0117] FIG. **10**D depicts a camera mount circuit board insert **1030**. This board is 29.9 mm wide and is press fit into the housing **1020** of the camera mount, shown alongside. The smaller hole **1031** of the two circular holes is a hole for the optical path. In some embodiments, a pin hole or fish eye lens is attached to this opening. The remaining white regions are where an electrical conductor, such as copper, is deposited. Dark regions are insulated. The 5 squares **1032** at the bottom of the image (near the scale bar **1039**) accommodate a switch. The central circle **1033** and circuitry are for 2 stacked 3 Volt watch batteries, and the 16 squares (pads **1034**) toward the top of the image are where the LEDs and resistors are soldered. After the holes are cut and the board **1030** is press fit into the housing **1020**, the now combined housing and board is a camera mount that can be affixed to any camera or camera and lens system using double sided tape or double sided foam tape.

[0118] FIG. **10**E depicts manufacturing drawings for the mouthpiece **1010**, including an aligned top view **1010***a*, front view **1010***b*, and side view **1010***c*, for a particular embodiment. In this embodiment, the distance associated with various dimensions are given in millimeters. Thus, the mouthpiece width is about 60.17 mm, height is about 41.91 mm. Each bite guide **1012** is about 41.07 mm wide, about 1.65 mm high and separated by about 31.75 mm. The mouthpiece is curved with a radius of curvature of about 38.1 mm; and is about 1.59 mm thick. Opening **1011** is about 42.93 mm wide and about 25 mm high. The sides of the opening **1011** are curved with a radius of curvature of 50.42 mm.

4. Hardware Overview

[0119] FIG. **11** is a block diagram that illustrates a computer system **1100** upon which an embodiment of the invention may be implemented. Computer system **1100** includes a communication mechanism such as a bus **1110** for passing information between other internal and external components of the computer system **1100**. Information is represented as physical signals of a measurable phenomenon, typically electric voltages, but including, in other embodiments, such phenomena as magnetic, electromagnetic, pressure, chemical, molecular atomic and quantum interactions. For example, north and south magnetic fields, or a zero and non-zero electric voltage, represent two states (0, 1) of a binary digit (bit).). Other phenomena can represent digits of a higher base. A superposition of multiple simultaneous quantum states before measurement represents a quantum bit (qubit). A sequence of one or more digits constitutes digital data that is used to represent a number or code for a character. In some embodiments, information called analog data is represented by a near continuum of measurable values within a particular range. Computer system **1100**, or a portion thereof, constitutes a means for performing one or more steps of one or more methods described herein.

[0120] A sequence of binary digits constitutes digital data that is used to represent a number or code for a character. A bus **1110** includes many parallel conductors of information so that information is transferred quickly among devices coupled to the bus **1110**. One or more processors **1102** for processing information are coupled with the bus **1110**. A processor **1102** performs a set of operations on information. The set of operations include bringing information in from the bus **1110** and placing information on the bus **1110**. The set of operations also typically include comparing two or more units of information, shifting positions of units of information, and combining two or more units of information, such as by addition or multiplication. A sequence of operations to be executed by the processor **1102** constitute computer instructions.

[0121] Computer system **1100** also includes a memory **1104** coupled to bus **1110**. The memory **1104**, such as a random access memory (RAM) or other dynamic storage device, stores information including computer instructions. Dynamic memory allows information stored therein to be changed by the computer system **1100**. RAM allows a unit of information stored at a location called a memory address to be stored and retrieved independently of information at neighboring addresses. The memory **1104** is also used by the processor **1102** to store temporary values during execution of computer instructions. The computer system **1100** also includes a read only memory (ROM) **1106** or other static storage device coupled to the bus **1110** for storing static information, including instructions, that is not changed by the computer system **1100**. Also coupled to bus **1110** is a non-volatile (persistent) storage device **1108**, such as a magnetic disk or optical disk, for storing information, including instructions, that persists even when the computer system **1100** is turned off or otherwise loses power.

[0122] Information, including instructions, is provided to the bus **1110** for use by the processor from an external input device **1112**, such as a keyboard containing alphanumeric keys operated by a human user, or a sensor. A sensor detects conditions in its vicinity and transforms those detections into signals compatible with the signals used to represent information in computer system **1100**. Other external devices coupled to bus **1110**, used primarily for interacting with humans, include a display device **1114**, such as a cathode ray tube (CRT) or a liquid crystal display (LCD), for presenting images, and a pointing device **1116**, such as a mouse or a trackball or cursor direction keys, for controlling a position of a small cursor image presented on the display **1114** and issuing commands associated with graphical elements presented on the display **1114**.

[0123] In the illustrated embodiment, special purpose hardware, such as an application specific integrated circuit (IC) **1120**, is coupled to bus **1110**. The special purpose hardware is configured to perform operations not performed by processor **1102** quickly enough for special purposes. Examples of application specific ICs include graphics accelerator cards for generating images for display **1114**, cryptographic boards for encrypting and decrypting messages sent over a network, speech recognition, and interfaces to special external devices, such as robotic arms and medical scanning equipment that repeatedly perform some complex sequence of operations that are more efficiently implemented in hardware.

[0124] Computer system **1100** also includes one or more instances of a communications interface **1170** coupled to bus **1110**. Communication interface **1170** provides a two-way communication coupling to a variety of external devices that operate with their own processors, such as printers, scanners and external disks. In general the coupling is with a network link **1178** that is connected to a local network **1180** to which

a variety of external devices with their own processors are connected. For example, communication interface **1170** may be a parallel port or a serial port or a universal serial bus (USB) port on a personal computer. In some embodiments, communications interface **1170** is an integrated services digital network (ISDN) card or a digital subscriber line (DSL) card or a telephone modem that provides an information communication connection to a corresponding type of telephone line. In some embodiments, a communication interface **1170** is a cable modem that converts signals on bus **1110** into signals for a communication connection over a coaxial cable or into optical signals for a communication connection over a fiber optic cable. As another example, communications interface **1170** may be a local area network (LAN) card to provide a data communication connection to a compatible LAN, such as Ethernet. Wireless links may also be implemented. Carrier waves, such as acoustic waves and electromagnetic waves, including radio, optical and infrared waves travel through space without wires or cables. Signals include man-made variations in amplitude, frequency, phase, polarization or other physical properties of carrier waves. For wireless links, the communications interface **1170** sends and receives electrical, acoustic or electromagnetic signals, including infrared and optical signals, that carry information streams, such as digital data.

[0125] The term computer-readable medium is used herein to refer to any medium that participates in providing information to processor **1102**, including instructions for execution. Such a medium may take many forms, including, but not limited to, non-volatile media, volatile media and transmission media. Non-volatile media include, for example, optical or magnetic disks, such as storage device **1108**. Volatile media include, for example, dynamic memory **1104**. Transmission media include, for example, coaxial cables, copper wire, fiber optic cables, and waves that travel through space without wires or cables, such as acoustic waves and electromagnetic waves, including radio, optical and infrared waves. The term computer-readable storage medium is used herein to refer to any medium that participates in providing information to processor **1102**, except for transmission media.

[0126] Common forms of computer-readable media include, for example, a floppy disk, a flexible disk, a hard disk, a magnetic tape, or any other magnetic medium, a compact disk ROM (CD-ROM), a digital video disk (DVD) or any other optical medium, punch cards, paper tape, or any other physical medium with patterns of holes, a RAM, a programmable ROM (PROM), an erasable PROM (EPROM), a FLASH-EPROM, or any other memory chip or cartridge, a carrier wave, or any other medium from which a computer can read. The term non-transitory computer-readable storage medium is used herein to refer to any medium that participates in providing information to processor **1102**, except for carrier waves and other signals.

[0127] Logic encoded in one or more tangible media includes one or both of processor instructions on a computer-readable storage media and special purpose hardware, such as ASIC **1120**.

[0128] Network link **1178** typically provides information communication through one or more networks to other devices that use or process the information. For example, network link **1178** may provide a connection through local network **1180** to a host computer **1182** or to equipment **1184** operated by an Internet Service Provider (ISP). ISP equipment **1184** in turn provides data communication services through the public, world-wide packet-switching communication network of networks now commonly referred to as the Internet **1190**. A computer called a server **1192** connected to the Internet provides a service in response to information received over the Internet. For example, server **1192** provides information representing video data for presentation at display **1114**.

[0129] The invention is related to the use of computer system **1100** for implementing the techniques described herein. According to one embodiment of the invention, those techniques are performed by computer system **1100** in response to processor **1102** executing one or more sequences of one or more instructions contained in memory **1104**. Such instructions, also called software and program code, may be read into memory **1104** from another computer-readable medium such as storage device **1108**. Execution of the sequences of instructions contained in memory **1104** causes processor **1102** to perform the method steps described herein. In alternative embodiments, hardware, such as application specific integrated circuit **1120**, may be used in place of or in combination with software to implement the invention. Thus, embodiments of the invention are not limited to any specific combination of hardware and software.

[0130] The signals transmitted over network link **1178** and other networks through communications interface **1170**, carry information to and from computer system **1100**. Computer system **1100** can send and receive information, including program code, through the networks **1180**, **1190** among others, through network link **1178** and communications interface **1170**. In an example using the Internet **1190**, a server **1192** transmits program code for a particular application, requested by a message sent from computer **1100**, through Internet **1190**, ISP equipment **1184**, local network **1180** and communications interface **1170**. The received code may be executed by processor **1102** as it is received, or may be stored in storage device **1108** or other non-volatile storage for later execution, or both. In this manner, computer system **1100** may obtain application program code in the form of a signal on a carrier wave.

[0131] Various forms of computer readable media may be involved in carrying one or more sequence of instructions or data or both to processor **1102** for execution. For example, instructions and data may initially be carried on a magnetic disk of a remote computer such as host **1182**. The remote computer loads the instructions and data into its dynamic memory and sends the instructions and data over a telephone line using a modem. A modem local to the computer system **1100** receives the instructions and data on a telephone line and uses an infra-red transmitter to convert the instructions and data to a signal on an infra-red a carrier wave serving as the network link **1178**. An infrared detector serving as communications interface **1170** receives the instructions and data carried in the infrared signal and places information representing the instructions and data onto bus **1110**. Bus **1110** carries the information to memory **1104** from which processor **1102** retrieves and executes the instructions using some of the data sent with the instructions. The instructions and data received in memory **1104** may optionally be stored on storage device **1108**, either before or after execution by the processor **1102**.

[0132] FIG. **12** illustrates a chip set **1200** upon which an embodiment of the invention may be implemented. Chip set **1200** is programmed to perform one or more steps of a method described herein and includes, for instance, the pro-

cessor and memory components described with respect to FIG. **11** incorporated in one or more physical packages (e.g., chips). By way of example, a physical package includes an arrangement of one or more materials, components, and/or wires on a structural assembly (e.g., a baseboard) to provide one or more characteristics such as physical strength, conservation of size, and/or limitation of electrical interaction. It is contemplated that in certain embodiments the chip set can be implemented in a single chip. Chip set **1200**, or a portion thereof, constitutes a means for performing one or more steps of a method described herein.

[0133] In one embodiment, the chip set **1200** includes a communication mechanism such as a bus **1201** for passing information among the components of the chip set **1200**. A processor **1203** has connectivity to the bus **1201** to execute instructions and process information stored in, for example, a memory **1205**. The processor **1203** may include one or more processing cores with each core configured to perform independently. A multi-core processor enables multiprocessing within a single physical package. Examples of a multi-core processor include two, four, eight, or greater numbers of processing cores. Alternatively or in addition, the processor **1203** may include one or more microprocessors configured in tandem via the bus **1201** to enable independent execution of instructions, pipelining, and multithreading. The processor **1203** may also be accompanied with one or more specialized components to perform certain processing functions and tasks such as one or more digital signal processors (DSP) **1207**, or one or more application-specific integrated circuits (ASIC) **1209**. A DSP **1207** typically is configured to process real-world signals (e.g., sound) in real time independently of the processor **1203**. Similarly, an ASIC **1209** can be configured to performed specialized functions not easily performed by a general purposed processor. Other specialized components to aid in performing the inventive functions described herein include one or more field programmable gate arrays (FPGA) (not shown), one or more controllers (not shown), or one or more other special-purpose computer chips.

[0134] The processor **1203** and accompanying components have connectivity to the memory **1205** via the bus **1201**. The memory **1205** includes both dynamic memory (e.g., RAM, magnetic disk, writable optical disk, etc.) and static memory (e.g., ROM, CD-ROM, etc.) for storing executable instructions that when executed perform one or more steps of a method described herein. The memory **1205** also stores the data associated with or generated by the execution of one or more steps of the methods described herein.

[0135] FIG. **13** is a diagram of exemplary components of a mobile terminal **1300** (e.g., cell phone handset) for communications, which is capable of operating in the system of FIG. **2**C, according to one embodiment. In some embodiments, mobile terminal **1301**, or a portion thereof, constitutes a means for performing one or more steps described herein. Generally, a radio receiver is often defined in terms of front-end and back-end characteristics. The front-end of the receiver encompasses all of the Radio Frequency (RF) circuitry whereas the back-end encompasses all of the base-band processing circuitry. As used in this application, the term "circuitry" refers to both: (1) hardware-only implementations (such as implementations in only analog and/or digital circuitry), and (2) to combinations of circuitry and software (and/or firmware) (such as, if applicable to the particular context, to a combination of processor(s), including digital signal processor(s), software, and memory(ies) that work together to cause an apparatus, such as a mobile phone or server, to perform various functions). This definition of "circuitry" applies to all uses of this term in this application, including in any claims. As a further example, as used in this application and if applicable to the particular context, the term "circuitry" would also cover an implementation of merely a processor (or multiple processors) and its (or their) accompanying software/or firmware. The term "circuitry" would also cover if applicable to the particular context, for example, a baseband integrated circuit or applications processor integrated circuit in a mobile phone or a similar integrated circuit in a cellular network device or other network devices.

[0136] Pertinent internal components of the telephone include a Main Control Unit (MCU) **1303**, a Digital Signal Processor (DSP) **1305**, and a receiver/transmitter unit including a microphone gain control unit and a speaker gain control unit. A main display unit **1307** provides a display to the user in support of various applications and mobile terminal functions that perform or support the steps as described herein. The display **1307** includes display circuitry configured to display at least a portion of a user interface of the mobile terminal (e.g., mobile telephone). Additionally, the display **1307** and display circuitry are configured to facilitate user control of at least some functions of the mobile terminal. An audio function circuitry **1309** includes a microphone **1311** and microphone amplifier that amplifies the speech signal output from the microphone **1311**. The amplified speech signal output from the microphone **1311** is fed to a coder/decoder (CODEC) **1313**.

[0137] A radio section **1315** amplifies power and converts frequency in order to communicate with a base station, which is included in a mobile communication system, via antenna **1317**. The power amplifier (PA) **1319** and the transmitter/modulation circuitry are operationally responsive to the MCU **1303**, with an output from the PA **1319** coupled to the duplexer **1321** or circulator or antenna switch, as known in the art. The PA **1319** also couples to a battery interface and power control unit **1320**.

[0138] In use, a user of mobile terminal **1301** speaks into the microphone **1311** and his or her voice along with any detected background noise is converted into an analog voltage. The analog voltage is then converted into a digital signal through the Analog to Digital Converter (ADC) **1323**. The control unit **1303** routes the digital signal into the DSP **1305** for processing therein, such as speech encoding, channel encoding, encrypting, and interleaving. In one embodiment, the processed voice signals are encoded, by units not separately shown, using a cellular transmission protocol such as enhanced data rates for global evolution (EDGE), general packet radio service (GPRS), global system for mobile communications (GSM), Internet protocol multimedia subsystem (IMS), universal mobile telecommunications system (UMTS), etc., as well as any other suitable wireless medium, e.g., microwave access (WiMAX), Long Term Evolution (LTE) networks, code division multiple access (CDMA), wideband code division multiple access (WCDMA), wireless fidelity (WiFi), satellite, and the like, or any combination thereof.

[0139] The encoded signals are then routed to an equalizer **1325** for compensation of any frequency-dependent impairments that occur during transmission though the air such as phase and amplitude distortion. After equalizing the bit stream, the modulator **1327** combines the signal with a RF

signal generated in the RF interface **1329**. The modulator **1327** generates a sine wave by way of frequency or phase modulation. In order to prepare the signal for transmission, an up-converter **1331** combines the sine wave output from the modulator **1327** with another sine wave generated by a synthesizer **1333** to achieve the desired frequency of transmission. The signal is then sent through a PA **1319** to increase the signal to an appropriate power level. In practical systems, the PA **1319** acts as a variable gain amplifier whose gain is controlled by the DSP **1305** from information received from a network base station. The signal is then filtered within the duplexer **1321** and optionally sent to an antenna coupler **1335** to match impedances to provide maximum power transfer. Finally, the signal is transmitted via antenna **1317** to a local base station. An automatic gain control (AGC) can be supplied to control the gain of the final stages of the receiver. The signals may be forwarded from there to a remote telephone which may be another cellular telephone, any other mobile phone or a land-line connected to a Public Switched Telephone Network (PSTN), or other telephony networks.

[0140] Voice signals transmitted to the mobile terminal **1301** are received via antenna **1317** and immediately amplified by a low noise amplifier (LNA) **1337**. A down-converter **1339** lowers the carrier frequency while the demodulator **1341** strips away the RF leaving only a digital bit stream. The signal then goes through the equalizer **1325** and is processed by the DSP **1305**. A Digital to Analog Converter (DAC) **1343** converts the signal and the resulting output is transmitted to the user through the speaker **1345**, all under control of a Main Control Unit (MCU) **1303** which can be implemented as a Central Processing Unit (CPU) (not shown).

[0141] The MCU **1303** receives various signals including input signals from the keyboard **1347**. The keyboard **1347** and/or the MCU **1303** in combination with other user input components (e.g., the microphone **1311**) comprise a user interface circuitry for managing user input. The MCU **1303** runs a user interface software to facilitate user control of at least some functions of the mobile terminal **1301** as described herein. The MCU **1303** also delivers a display command and a switch command to the display **1307** and to the speech output switching controller, respectively. Further, the MCU **1303** exchanges information with the DSP **1305** and can access an optionally incorporated SIM card **1349** and a memory **1351**. In addition, the MCU **1303** executes various control functions required of the terminal. The DSP **1305** may, depending upon the implementation, perform any of a variety of conventional digital processing functions on the voice signals. Additionally, DSP **1305** determines the background noise level of the local environment from the signals detected by microphone **1311** and sets the gain of microphone **1311** to a level selected to compensate for the natural tendency of the user of the mobile terminal **1301**.

[0142] The CODEC **1313** includes the ADC **1323** and DAC **1343**. The memory **1351** stores various data including call incoming tone data and is capable of storing other data including music data received via, e.g., the global Internet. The software module could reside in RAM memory, flash memory, registers, or any other form of writable storage medium known in the art. The memory device **1351** may be, but not limited to, a single memory, CD, DVD, ROM, RAM, EEPROM, optical storage, magnetic disk storage, flash memory storage, or any other non-volatile storage medium capable of storing digital data.

[0143] An optionally incorporated SIM card **1349** carries, for instance, important information, such as the cellular phone number, the carrier supplying service, subscription details, and security information. The SIM card **1349** serves primarily to identify the mobile terminal **1301** on a radio network. The card **1349** also contains a memory for storing a personal telephone number registry, text messages, and user specific mobile terminal settings.

[0144] In some embodiments, the mobile terminal **1301** includes a digital camera comprising an array of optical detectors, such as charge coupled device (CCD) array **1365**. The output of the array is image data that is transferred to the MCU for further processing or storage in the memory **1351** or both. In the illustrated embodiment, the light impinges on the optical array through a lens **1363**, such as a pin-hole lens or a material lens made of an optical grade glass or plastic material. In the illustrated embodiment, the mobile terminal **1301** includes a light source **1361**, such as a LED to illuminate a subject for capture by the optical array, e.g., CCD **1365**. The light source is powered by the battery interface and power control module **1320** and controlled by the MCU **1303** based on instructions stored or loaded into the MCU **1303**.

5. Extensions and Alternatives

[0145] In the foregoing specification, the invention has been described with reference to specific embodiments thereof. It will, however, be evident that various modifications and changes may be made thereto without departing from the broader spirit and scope of the invention. The specification and drawings are, accordingly, to be regarded in an illustrative rather than a restrictive sense. Throughout this specification and the claims, unless the context requires otherwise, the word "comprise" and its variations, such as "comprises" and "comprising," will be understood to imply the inclusion of a stated item, element or step or group of items, elements or steps but not the exclusion of any other item, element or step or group of items. elements or steps. Furthermore, the indefinite article "a" or "an" is meant to indicate one or more of the item, element or step modified by the article.

What is claimed is:

1. An oral cavity image bracket comprising:

a mouthpiece of rigid material comprising an upper bite guide and a lower bite guide, both disposed on a posterior side of the mouthpiece and separated by an opening through the mouthpiece, wherein the upper bite guide and lower bite guide are spaced apart such that a subject biting down on the upper bite guide with the subject's upper jaw and biting up on the lower bite guide with the subject's lower jaw opens the subject's oral cavity to inspection through the opening; and

a camera mount disposed on an anterior side of the mouthpiece and comprising a flange configured to engage and slide along the opening of the mouthpiece,

wherein the camera mount further comprises

an optical path configured for light to pass through the camera mount and through the opening in the mouthpiece, and

a clip disposed on an anterior side of the camera mount, wherein the clip is configured to removeably hold a camera on the anterior side of the camera mount to record light passing through the optical path from the posterior side of the camera mount.

2. An oral cavity image bracket as recited in claim **1**, wherein the clip is configured to removeabley hold a camera

selected from a group comprising: a digital camera; a digital camera with on board processor; a cell phone with digital camera; a programmable cell phone with digital camera.

3. An oral cavity image bracket as recited in claim **1**, wherein the optical path comprises a pin hole lens.

4. An oral cavity image bracket as recited in claim **1**, wherein the optical path comprises a material lens.

5. An oral cavity image bracket as recited in claim **1**, wherein the optical path comprises a removeable optical filter that blocks light from a light source and passes fluorescent light emitted by tissue in the oral cavity of the subject in response to the light source.

6. An oral cavity image bracket as recited in claim **1**, wherein the mouthpiece further comprises a light source disposed on the posterior side of the mouthpiece and configured to illuminate the oral cavity of the subject.

7. An oral cavity image bracket as recited in claim **6**, wherein at least one of the mouthpiece or the camera mount comprises a power source configured to supply power to the light source.

8. An oral cavity image system comprising:

a bracket comprising

a mouthpiece of rigid material comprising an upper bite guide and a lower bite guide, both disposed on a posterior side of the mouthpiece and separated by an opening through the mouthpiece, wherein the upper bite guide and lower bite guide are spaced apart such that a subject biting down on the upper bite guide with the subject's upper jaw and biting up on the lower bite guide with the subject's lower jaw opens the subject's oral cavity to inspection through the opening; and

a camera mount disposed on an anterior side of the mouthpiece and comprising a flange configured to engage and slide along the opening of the mouthpiece,

wherein the camera mount further comprises an optical path for light to pass through the camera mount and through the opening in the mouthpiece;

a camera removeably attached to the anterior side of the camera mount, wherein the camera is configured to record and display an image based on light passing through the optical path from the posterior side of the camera mount; and

a processor configured to merge data from a plurality of images recorded by the camera at a corresponding plurality of positions of the camera mount as the camera mount slides along the opening in the mouthpiece.

9. An oral cavity image system as recited in claim **8**, wherein the posterior side of the mouthpiece further comprises a light source configured to illuminate the oral cavity of the subject.

10. An oral cavity image system as recited in claim **8**, wherein the camera is selected from a group comprising: a digital camera; a digital camera with the processor on board; a cell phone with digital camera; a programmable cell phone with digital camera and with the processor on board.

11. An oral cavity image system as recited in claim **8**, wherein the camera further comprises a light source configured to illuminate the oral cavity of the subject.

12. An oral cavity image system as recited in claim **8**, wherein

the system further comprises a communications module configured to communicate with a remote server; and

the processor includes a processor on the remote server.

13. An oral cavity image system as recited in claim **8**, wherein the camera is further configured to display an image based on the data merged by the processor from a plurality of images recorded by the camera.

14. A method comprising:

removeably attaching a camera to a camera mount of a bracket comprising the camera mount and a mouthpiece, wherein

the mouthpiece comprises an upper bite guide and a lower bite guide, both disposed on a posterior side of the mouthpiece and separated by an opening through the mouthpiece, wherein the upper bite guide and lower bite guide are spaced apart such that a subject biting down on the upper bite guide with the subject's upper jaw and biting up on the lower bite guide with the subject's lower jaw opens the subject's oral cavity to inspection through the opening,

the camera mount is disposed on an anterior side of the mouthpiece and comprises a flange configured to engage and slide along the opening of the mouthpiece and an optical path for light to pass through the camera mount and through the opening in the mouthpiece, and

the camera is configured to record and display an image based on light passing through the optical path from the posterior side of the camera mount;

causing a subject to bite against the bite guides of the mouthpiece;

sliding the camera mount to a plurality of positions along the opening in the mouthpiece; and

causing the camera to capture a plurality of images corresponding to the plurality of positions.

15. A method as recited in claim **14**, further comprising using a processor to merge data from the plurality of images into a standard image.

16. A method as recited in claim **15**, further comprising displaying the standard image.

17. A method as recited in claim **16**, further comprising viewing the standard image and determining a condition of subject based on the standard image.

18. A method as recited in claim **14**, further comprising inserting a filter into the optical path, wherein the filter blocks light from a light source that illuminates the subject's oral cavity and passes light fluorescently emitted by tissue in the subject's oral cavity

19. A method as recited in claim **14**, further comprising activating a light source to illuminate the subject's oral cavity.

20. A method as recited in claim **14**, wherein the light source is disposed on an posterior surface of the mouthpiece.

21. A method as recited in claim **14**, wherein the camera is a programmable cell phone with camera and on-board processor.

22. An apparatus comprising:

at least one processor; and

at least one memory including one or more sequences of instructions,

the at least one memory and the one or more sequences of instructions configured to, with the at least one processor, cause the apparatus to perform at least the following,

determining distance from an imaging plane to a surface of an oral cavity of a subject based on relative intensity of a pixel in an image frame captured at the imaging plane compared to adjacent pixels in the image frame and a model of focusing optics for each

of a plurality of image frames of the oral cavity corresponding to a plurality of different look directions; and

merging, into a single image, pixels from the plurality of image frames of the oral cavity.

23. A non-transitory computer-readable medium carrying one or more sequences of instructions, wherein execution of the one or more sequences of instructions by one or more processors causes an apparatus to perform the steps of:

determining distance from an imaging plane to a surface of an oral cavity of a subject based on relative intensity of a pixel in an image frame captured at the imaging plane compared to adjacent pixels in the image frame and a model of focusing optics for each of a plurality of image frames of the oral cavity corresponding to a plurality of different look directions; and

merging, into a single image, pixels from the plurality of image frames of the oral cavity.

* * * * *

EXHIBIT R-9

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
16/951,401	11/18/2020	Philippe SALAH	N&P-51400US2	3930

108676 7590 10/22/2021

Ronald M. Kachmarik
Cooper Legal Group LLC
1388 Ridge Road, Unit 1
Hinckley, OH 44233

EXAMINER
HASAN, MAINUL

ART UNIT	PAPER NUMBER
2485	

NOTIFICATION DATE	DELIVERY MODE
10/22/2021	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

docketing@cooperlegalgroup.com

Office Action Summary	Application No. 16/951,401	Applicant(s) SALAH et al.	
	Examiner MAINUL HASAN	Art Unit 2485	AIA (FITF) Status Yes

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTHS FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☑ Responsive to communication(s) filed on 15 September 2021.

☐ A declaration(s)/affidavit(s) under **37 CFR 1.130(b)** was/were filed on _____.

2a)☐ This action is **FINAL.** 2b) ☑ This action is non-final.

3)☐ An election was made by the applicant in response to a restriction requirement set forth during the interview on _____; the restriction requirement and election have been incorporated into this action.

4)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims*

5) ☑ Claim(s) 1-28 is/are pending in the application.

5a) Of the above claim(s) _____ is/are withdrawn from consideration.

6) ☐ Claim(s) _____ is/are allowed.

7) ☑ Claim(s) 1-28 is/are rejected.

8) ☐ Claim(s) _____ is/are objected to.

9) ☐ Claim(s) _____ are subject to restriction and/or election requirement

* If any claims have been determined allowable, you may be eligible to benefit from the **Patent Prosecution Highway** program at a participating intellectual property office for the corresponding application. For more information, please see http://www.uspto.gov/patents/init_events/pph/index.jsp or send an inquiry to **PPHfeedback@uspto.gov.**

Application Papers

10)☑ The specification is objected to by the Examiner.

11)☑ The drawing(s) filed on 16 March 2021 is/are: a)☑ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

Priority under 35 U.S.C. § 119

12)☑ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

Certified copies:

a)☑ All b)☐ Some** c)☐ None of the:

1.☑ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

** See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☑ Notice of References Cited (PTO-892)

2) ☑ Information Disclosure Statement(s) (PTO/SB/08a and/or PTO/SB/08b) Paper No(s)/Mail Date _____.

3) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date _____.

4) ☐ Other: _____.

DETAILED ACTION

Notice of Pre-AIA or AIA Status

The present application, filed on or after March 16, 2013, is being examined under the first inventor to file provisions of the AIA. In the event the determination of the status of the application as subject to AIA 35 U.S.C. 102 and 103 (or as subject to pre-AIA 35 U.S.C. 102 and 103) is incorrect, any correction of the statutory basis for the rejection will not be considered a new ground of rejection if the prior art relied upon, and the rationale supporting the rejection, would be the same under either status. There are a total of 28 claims and claims 1-28 are pending.

Examiner's Note

The Examiner noted that the Applicant has identified the alleged claim and specification filed on 03/16/2021 as Exhibit A and not amended claim and specification. The drawing filed on 03/16/2021 is the amended drawing. The Examiner agrees with the Applicant's explanation and therefore, the current Office Action is based on the claim set filed on 11/18/2020. Since the previous Office Action dated 05/13/2021 has been withdrawn, this is the first Non-Final Office Action for this application and has no bearing on any previous Office Actions.

Specification

Applicant is reminded of the proper content of an abstract of the disclosure.

A patent abstract is a concise statement of the technical disclosure of the patent and should include that which is new in the art to which the invention pertains. The abstract should

not refer to purported merits or speculative applications of the invention and should not compare the invention with the prior art.

If the patent is of a basic nature, the entire technical disclosure may be new in the art, and the abstract should be directed to the entire disclosure. If the patent is in the nature of an improvement in an old apparatus, process, product, or composition, the abstract should include the technical disclosure of the improvement. The abstract should also mention by way of example any preferred modifications or alternatives.

Where applicable, the abstract should include the following: (1) if a machine or apparatus, its organization and operation; (2) if an article, its method of making; (3) if a chemical compound, its identity and use; (4) if a mixture, its ingredients; (5) if a process, the steps.

Extensive mechanical and design details of an apparatus should not be included in the abstract. **The abstract should be in narrative form and generally limited to a single paragraph within the range of 50 to 150 words in length**.

See MPEP § 608.01(b) for guidelines for the preparation of patent abstracts.

The abstract filed on 03/16/2021 does not follow the guideline as underlined in bold text above. Appropriate action is required.

Double Patenting

The nonstatutory double patenting rejection is based on a judicially created doctrine grounded in public policy (a policy reflected in the statute) so as to prevent the unjustified or improper timewise extension of the "right to exclude" granted by a patent and to prevent possible harassment by multiple assignees. A nonstatutory double patenting rejection is appropriate where

the conflicting claims are not identical, but at least one examined application claim is not patentably distinct from the reference claim(s) because the examined application claim is either anticipated by, or would have been obvious over, the reference claim(s). See, e.g., *In re Berg*, 140 F.3d 1428, 46 USPQ2d 1226 (Fed. Cir. 1998); *In re Goodman*, 11 F.3d 1046, 29 USPQ2d 2010 (Fed. Cir. 1993); *In re Longi*, 759 F.2d 887, 225 USPQ 645 (Fed. Cir. 1985); *In re Van Ornum*, 686 F.2d 937, 214 USPQ 761 (CCPA 1982); *In re Vogel*, 422 F.2d 438, 164 USPQ 619 (CCPA 1970); *In re Thorington*, 418 F.2d 528, 163 USPQ 644 (CCPA 1969).

A timely filed terminal disclaimer in compliance with 37 CFR 1.321(c) or 1.321(d) may be used to overcome an actual or provisional rejection based on nonstatutory double patenting provided the reference application or patent either is shown to be commonly owned with the examined application, or claims an invention made as a result of activities undertaken within the scope of a joint research agreement. See MPEP § 717.02 for applications subject to examination under the first inventor to file provisions of the AIA as explained in MPEP § 2159. See MPEP § 2146 *et seq.* for applications not subject to examination under the first inventor to file provisions of the AIA. A terminal disclaimer must be signed in compliance with 37 CFR 1.321(b).

The USPTO Internet website contains terminal disclaimer forms, which may be used. Please visit www.uspto.gov/patent/patents-forms. The filing date of the application in which the form is filed determines what form (e.g., PTO/SB/25, PTO/SB/26, PTO/AIA/25, or PTO/AIA/26) should be used. A web-based eTerminal Disclaimer may be filled out completely online using web-screens. An eTerminal Disclaimer that meets all requirements is auto-processed and approved immediately upon submission. For more information about eTerminal Disclaimers, refer to www.uspto.gov/patents/process/file/efs/guidance/eTD-info-I.jsp.

Claims 1-28 of the instant application are rejected on the ground of nonstatutory obviousness type double patenting as being unpatentable over claims 1-17 of U.S. Patent No. **10,736,715 B2** in view of **Prakash et al. (US PGPub 2013/0209954 A1)**.

Claim **1** of the instant application is rejected on the ground of nonstatutory obviousness type double patenting as being unpatentable over claim 1 of U.S. Patent No. **10,736,715 B2** in view of **Prakash et al. (US PGPub 2013/0209954 A1)**. Although the claims at issue are not identical, the instant application claim is not patentable over the Patent claim in view of **Prakash et al.** as shown in the following table:

	16951401 (Instant Application)	**10,736,715 B2** (Patent)
	Claim 1	**Claim 1**
1	*A patient-operated* ***imaging device comprising:***	***An imaging device including:***
2	***- a support;***	***a support;***
3	***- a mouth retractor formed as an integral part of the support and defining a retractor opening; and***	***a mouth retractor fastened to the support and defining a retractor opening; and***
4	***- a mechanism for fastening an image acquisition apparatus to the support in a position in which the acquisition apparatus is oriented so as to receive an image of the retractor opening,***	***a mechanism for fastening an image acquisition apparatus to the support in a position in which the acquisition apparatus is oriented so as to receive an image of the retractor opening,***
5	***wherein the support takes the form of a box that is in communication with the outside via the retractor opening and via an acquisition opening through which the image acquisition apparatus fastened to the support receives the image of the retractor opening,***	***wherein the support takes the form of a box that is in communication with the outside via the retractor opening and via an acquisition opening through which the acquisition apparatus fastened to the support receives said image of the retractor opening,***
6	***the support being configured so that the image acquisition apparatus observes the retractor opening regardless of the configuration of the support;***	***the support being configured so that the acquisition apparatus observes the retractor opening regardless of a configuration of said support,***

7	*wherein the patient-operated imaging device is adapted to obtain a plurality of images, wherein at least two of the plurality of images correspond to different angles with respect to the patient's teeth.*	*said mechanism being chosen from the group consisting of clip-fastening means, self-gripping strips of hook and loop fastener type, clamping jaws, screws, magnets, and complementarity of shape between the support and the acquisition apparatus, or consisting of a cover that may be clamped against the support.*

The equivalence in claim limitations of the instant application and the patent are highlighted in **bold** texts. Although the instant application claim limitations appear to be a broader version of the corresponding Patent claim limitation, however, a close review shows that the limitations are identical except an additional limitation (#7) in the instant application, which is not present in the patent. However, **Prakash et al.** teach a system in the same field of endeavor **(Abstract)**, where it teaches limitation #7 of the instant application **(Prakash et al.; [0069]-[0070]; Figs. 3A-C** show capturing of plurality of images by the cellphone and as the drawings indicate the plurality of images are taken at different angles with respect to the patient's teeth). It would have been obvious before the effective filing date of the claimed invention to a person having ordinary skill in the art to combine the Patent's invention of a dental imaging device to include **Prakash et al**'s capturing of plurality of images from different angle with respect to the patient's teeth, because it allows merging of multiple raw images to allow for obtaining a wider field of view which can help get peripheral features (landscape mode) and in portrait mode, it enables a broader amount of the hard palate and floor of the mouth (area under tongue) **(Prakash et al.; [0072])**. Therefore, the instant application claim 1 as a whole is not patentable over the patent claim 1 in view of **Prakash et al.** This is a non-statutory obviousness type double patenting rejection.

Claim **28** of the instant application is rejected on the ground of nonstatutory obviousness type double patenting as being unpatentable over claim 1 of U.S. Patent No. **10,736,715 B2** in view of **Prakash et al. (US PGPub 2013/0209954 A1)**. Although the claims at issue are not identical, the instant application claim is not patentable over the Patent claim in view of **Prakash et al.** as shown in the following table:

	16951401 (Instant Application)	**10,736,715 B2** (Patent)
	Claim 28	**Claim 1**
1	*A patient-operated* ***imaging device comprising:***	***An imaging device including:***
2	***- a support;***	***a support;***
3	***- a mouth retractor formed as an integral part of the support and defining a retractor opening; and***	***a mouth retractor fastened to the support and defining a retractor opening; and***
4	***- an image acquisition apparatus fastened to the support in a position in which the image acquisition apparatus is oriented so as to receive an image of the retractor opening,***	***a mechanism for fastening an image acquisition apparatus to the support in a position in which the acquisition apparatus is oriented so as to receive an image of the retractor opening,***
5	***wherein the support takes the form of a box that is in communication with the outside via the retractor opening and via an acquisition opening through which the image acquisition apparatus fastened to the support receives the image of the retractor opening,***	***wherein the support takes the form of a box that is in communication with the outside via the retractor opening and via an acquisition opening through which the acquisition apparatus fastened to the support receives said image of the retractor opening,***
6	***the support being configured so that the image acquisition apparatus observes the retractor opening regardless of the configuration of the support,***	***the support being configured so that the acquisition apparatus observes the retractor opening regardless of a configuration of said support,***
7	*wherein the patient-operated imaging device is adapted to obtain a plurality of images, wherein at least two of the plurality of images correspond to different angles with respect to the patient's teeth.*	*said mechanism being chosen from the group consisting of clip-fastening means, self-gripping strips of hook and loop fastener type, clamping jaws, screws, magnets, and complementarity of shape between the support and the acquisition apparatus, or consisting of a cover that may be clamped against the support.*

The equivalence in claim limitations of the instant application and the patent are highlighted in **bold** texts. Although the instant application claim limitations appear to be a broader version of the corresponding Patent claim limitation, however, a close review shows that the limitations are identical except an additional limitation (#7) in the instant application, which is not present in the patent. However, **Prakash et al.** teach a system in the same field of endeavor **(Abstract)**, where it teaches limitation #7 of the instant application **(Prakash et al.; [0069]-[0070]; Figs. 3A-C** show capturing of plurality of images by the cellphone and as the drawings indicate the plurality of images are taken at different angles with respect to the patient's teeth). It would have been obvious before the effective filing date of the claimed invention to a person having ordinary skill in the art to combine the Patent's invention of a dental imaging device to include **Prakash et al**'s capturing of plurality of images from different angle with respect to the patient's teeth, because it allows merging of multiple raw images to allow for obtaining a wider field of view which can help get peripheral features (landscape mode) and in portrait mode, it enables a broader amount of the hard palate and floor of the mouth (area under tongue) (**Prakash et al.; [0072]**). Therefore, the instant application claim 28 as a whole is not patentable over the patent claim 1 in view of **Prakash et al.** This is a non-statutory obviousness type double patenting rejection.

Claims 2-27 of the instant application are rejected on the ground of nonstatutory obviousness type double patenting as being unpatentable over a combination of claims of U.S. Patent No. **10,736,715 B2** in view of **Prakash et al. (US PGPub 2013/0209954 A1)**.

Claims 1-28 of the instant application are also rejected on the ground of nonstatutory obviousness type double patenting as being unpatentable over claims 1-21 of U.S. Patent No. **10,842,592 B2** in view of **Prakash et al. (US PGPub 2013/0209954 A1)**.

Claim **1** of the instant application is rejected on the ground of nonstatutory obviousness type double patenting as being unpatentable over claim 1 of U.S. Patent No. **10,842,592 B2** in view of **Prakash et al. (US PGPub 2013/0209954 A1)**. Although the claims at issue are not identical, the instant application claim is not patentable over the Patent claim in view of **Prakash et al.** as shown in the following table:

	16951401 (Instant Application)	**10,842,592 B2** (Patent)
	Claim 1	**Claim 1**
1	*A patient-operated* ***imaging device comprising:***	***An imaging device including:***
2	***- a support;***	***a support;***
3	***- a mouth retractor formed as an integral part of the support and defining a retractor opening; and***	***a mouth retractor fastened, to the support and defining a retractor opening****, the mouth retractor including a rim extending around the retractor opening; a colorimetric calibration chart and/or a translucence calibration chart; and a light source that is oriented so as to illuminate both teeth of a patient through the retractor opening and said colorimetric calibration chart and/or said translucence calibration chart;*
4	*- a* ***mechanism for fastening an image acquisition apparatus to the support in a position in which the acquisition apparatus is oriented so as to receive an image of the retractor opening,***	***means for fastening an image acquisition apparatus to the support in a position in which the image acquisition apparatus is oriented so as to receive an image*** *of both* ***the retractor opening*** *and of said colorimetric calibration chart and/or of said translucence calibration chart,*

5	***wherein the support takes the form of a box that is in communication with the outside via the retractor opening and via an acquisition opening through which the image acquisition apparatus fastened to the support receives the image of the retractor opening,***	*wherein said means for fastening can be deactivated, the mouth retractor being configured so as, in a service position in which the mouth retractor is positioned on the mouth of a patient, lips of the patient may rest on said rim so that only an inside of the mouth is visible through the retractor opening;*
6	***the support being configured so that the image acquisition apparatus observes the retractor opening regardless of the configuration of the support;***	***the support taking the form of a box, that is in communication with the outside via the retractor opening and via an acquisition opening through which the acquisition apparatus fastened to the support receives said image of the retractor opening,***
7	*wherein the patient-operated imaging device is adapted to obtain a plurality of images, wherein at least two of the plurality of images correspond to different angles with respect to the patient's teeth.*	***the support being configured so that the acquisition apparatus observes the retractor opening regardless of a configuration of said support****; the image acquisition apparatus being a mobile phone or a tablet.*

The equivalence in claim limitations of the instant application and the patent are highlighted in **bold** texts. Although the instant application claim limitations appear to be a broader version of the corresponding Patent claim limitation, however, a close review shows that the limitations are identical except an additional limitation (#7) in the instant application, which is not present in the patent. However, **Prakash et al.** teach a system in the same field of endeavor **(Abstract)**, where it teaches limitation #7 of the instant application **(Prakash et al.; [0069]-[0070]; Figs. 3A-C** show capturing of plurality of images by the cellphone and as the drawings indicate the plurality of images are taken at different angles with respect to the patient's teeth). It would have been obvious before the effective filing date of the claimed invention to a person having ordinary skill in the art to combine the Patent's invention of a dental

imaging device to include **Prakash et al**'s capturing of plurality of images from different angle with respect to the patient's teeth, because it allows merging of multiple raw images to allow for obtaining a wider field of view which can help get peripheral features (landscape mode) and in portrait mode, it enables a broader amount of the hard palate and floor of the mouth (area under tongue) (**Prakash et al.; [0072]**). Therefore, the instant application claim 1 as a whole is not patentable over the patent claim 1 in view of **Prakash et al.** This is a non-statutory obviousness type double patenting rejection.

Claim **28** of the instant application is rejected on the ground of nonstatutory obviousness type double patenting as being unpatentable over claim 21 of U.S. Patent No. **10,842,592 B2** in view of **Prakash et al. (US PGPub 2013/0209954 A1)**. Although the claims at issue are not identical, the instant application claim is not patentable over the Patent claim in view of **Prakash et al.** as shown in the following table:

	16951401 (Instant Application)	**10,842,592 B2** (Patent)
	Claim 28	**Claim 21**
1	*A patient-operated* ***imaging device comprising:***	***An imaging device including:***
2	***- a support;***	***a support;***
3	***- a mouth retractor formed as an integral part of the support and defining a retractor opening; and***	***a mouth retractor fastened, to the support and defining a retractor opening;***
4	***- an image acquisition apparatus fastened to the support in a position in which the image acquisition apparatus is oriented so as to receive an image of the retractor opening,***	*a colorimetric calibration chart and/or a translucence calibration chart; and a light source that is oriented so as to illuminate both teeth of the patient through the retractor opening and said colorimetric calibration chart and/or said translucence calibration chart;*
5	***wherein the support takes the form of a box that is in communication with the outside via the retractor opening and via an acquisition opening through which***	***means for fastening an image acquisition apparatus to the support in a position in which the acquisition apparatus is oriented so as to receive an image of both***

	the image acquisition apparatus fastened to the support receives the image of the retractor opening,	***the retractor opening*** *and of said colorimetric calibration chart and/or of said translucence calibration chart,*
6	***the support being configured so that the image acquisition apparatus observes the retractor opening regardless of the configuration of the support,***	*wherein said means for fastening can be deactivated, and wherein the light source is configured so as to project a reference frame toward the retractor opening; wherein the image acquisition apparatus is a mobile phone or a tablet;* ***wherein the support takes the form of a box, that is in communication with the outside via the retractor opening and via an acquisition opening through which the acquisition apparatus fastened to the support receives said image of the retractor opening****,*
7	*wherein the patient-operated imaging device is adapted to obtain a plurality of images, wherein at least two of the plurality of images correspond to different angles with respect to the patient's teeth.*	***the support being configured so that the acquisition apparatus observes the retractor opening regardless of a configuration of said support****.*

The equivalence in claim limitations of the instant application and the patent are highlighted in **bold** texts. Although the instant application claim limitations appear to be a broader version of the corresponding Patent claim limitation, however, a close review shows that the limitations are identical except an additional limitation (#7) in the instant application, which is not present in the patent. However, **Prakash et al.** teach a system in the same field of endeavor **(Abstract)**, where it teaches limitation #7 of the instant application **(Prakash et al.; [0069]-[0070]; Figs. 3A-C** show capturing of plurality of images by the cellphone and as the drawings indicate the plurality of images are taken at different angles with respect to the patient's teeth). It would have been obvious before the effective filing date of the claimed

invention to a person having ordinary skill in the art to combine the Patent's invention of a dental imaging device to include **Prakash et al**'s capturing of plurality of images from different angle with respect to the patient's teeth, because it allows merging of multiple raw images to allow for obtaining a wider field of view which can help get peripheral features (landscape mode) and in portrait mode, it enables a broader amount of the hard palate and floor of the mouth (area under tongue) (**Prakash et al.; [0072]**). Therefore, the instant application claim 28 as a whole is not patentable over the patent claim 21 in view of **Prakash et al.** This is a non-statutory obviousness type double patenting rejection.

Claims **2-27** of the instant application are also rejected on the ground of nonstatutory obviousness type double patenting as being unpatentable over a combination of claims of U.S. Patent No. **10,842,592 B2** in view of **Prakash et al. (US PGPub 2013/0209954 A1)**.

Claim Rejections - 35 USC § 112

The following is a quotation of 35 U.S.C. 112(b):

(B) CONCLUSION.—The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the inventor or a joint inventor regards as the invention.

The following is a quotation of 35 U.S.C. 112, second paragraph:

The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.

Claims 2-4, 10-27 are rejected under 35 U.S.C. 112(b) or 35 U.S.C. 112 (pre-AIA), second paragraph, as being indefinite for failing to particularly point out and distinctly claim the

subject matter which the inventor or a joint inventor (or for applications subject to pre-AIA 35 U.S.C. 112, the applicant), regards as the invention.

In claims 2-4, 10-27, reference is made of the element “the imaging device”. However, all the claims in question are directly or indirectly dependent on claim 1, where “a patient-operated imaging device” is recited. Therefore, the recitation of “the imaging device” creates an issue of insufficient antecedent basis for the particular element in claims 2-4, 10-27.

The term "*substantially*" in claim 14 is a relative term which renders the claim indefinite. The term " substantially" is not defined by the claim, the specification does not provide a standard for ascertaining the requisite degree, and one of ordinary skill in the art would not be reasonably apprised of the scope of the invention. Similar terms are used in claims 19 and 26.

Claim Rejections - 35 USC § 102

The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

> A person shall be entitled to a patent unless –
>
> (a)(1) the claimed invention was patented, described in a printed publication, or in public use, on sale or otherwise available to the public before the effective filing date of the claimed invention.

Claims 1, 3, 12-19, 21-28 are rejected under AIA 35 U.S.C. 102(a)(1) as being anticipated by Prakash et al. (US PGPub 2013/0209954 A1).

Regarding claim **1**, **Prakash et al.** disclose *a patient-operated imaging device* **(Fig. 5B)** *comprising:*

- *a support* **(Figs. 2A-C,** reference numeral **220. Figs. 5A-B,** reference numeral **520)**;

- *a mouth retractor formed as an integral part of the support and defining a retractor opening* **(Figs. 2A-B,** reference numeral **210** is the mouthpiece or retractor and **211** is the opening. **Figs. 5A-B,** reference numeral **510** is the mouthpiece or retractor and **511** is the opening. To be specific, reference numerals **212a-b**, known as bite guides, on the mouthpiece **210** act as the mouth retractors as shown in **Fig. 2A)**; *and*

- *a mechanism for fastening an image acquisition apparatus to the support in a position in which the acquisition apparatus is oriented so as to receive an image of the retractor opening* **([0062]; Fig. 2B,** reference numeral **226a-c**),

wherein the support takes the form of a box **(Fig. 5B,** reference numeral **520** shows the box shaped camera mount as the support) *that is in communication with the outside via the retractor opening and via an acquisition opening through which the image acquisition apparatus fastened to the support receives the image of the retractor opening* **([0062]; Fig. 5A** shows the retractor or the mouthpiece **510** and the opening **511** through which the image acquisition device or the cellphone camera receives the image for capturing), *the support being configured so that the image acquisition apparatus observes the retractor opening regardless of the configuration of the support* **([0069]-[0070]; Figs. 3A-C** show the cellphone camera acquiring the images of the oral cavity through the opening regardless of the configuration of the mount);

the mechanism being chosen from the group consisting of an elastic member, clip-fastening means **([0062]; Fig. 2B,** reference numeral **226a-c**), *self-gripping strips of hook and loop fastener type, clamping jaws, screws, magnets, and complementarity of shape between the support and the image acquisition apparatus, or consisting of a cover that may be clamped against the support,*

wherein the patient-operated imaging device is adapted to obtain a plurality of images, wherein at least two of the plurality of images correspond to different angles with respect to the patient's teeth **([0069]-[0070]; Figs. 3A-C** show capturing of plurality of images by the cellphone and as the drawings indicate the plurality of images are taken at different angles with respect to the patient's teeth).

Regarding claim **3**, **Prakash et al.** disclose *an imaging kit* **(Fig. 2C,** reference numeral **250)** *comprising:*

- the imaging device as claimed in claim 1 **(Fig. 2C,** reference numeral **200)***; and*

- an image acquisition apparatus **(Fig. 2C,** reference numeral **280)** *that is fastened to the device in a position in which the image acquisition apparatus is oriented to receive an image of the retractor opening* **([0062]; Fig. 2B,** reference numeral **226a-c)**.

Regarding claim **12**, **Prakash et al.** disclose *the imaging device as claimed in claim 1, in which the support defines a closed chamber when the opening of the mouth retractor and the acquisition opening are obturated* **(Fig. 5B,** reference numeral **520)**.

Regarding claim **13**, **Prakash et al.** disclose *the imaging device as claimed in claim 12, in which a lateral wall delimiting the chamber is formed or consists of a material that does not allow the content of the chamber to be accurately discerned* **(Fig. 5B,** reference numeral **520)**.

Regarding claim **14**, **Prakash et al.** disclose *the imaging device as claimed in claim 13, in which the lateral wall is opaque, so that an inner volume of the chamber receives substantially no light from outside of the chamber in a service position* **(Fig. 5B,** reference numeral **520)**.

Regarding claim **15**, **Prakash et al.** disclose *the imaging device as claimed in claim 1, in which the mouth retractor comprises lobes that are arranged so as to spread cheeks of a patient away from teeth of the patient* **(Fig. 2A,** reference numerals **212a-b** and **Fig. 5A,** reference numerals **512** represent the bite guides which act as a spreader of the patient's cheek as described in **[0004], L3-10**).

Regarding claim **16**, **Prakash et al.** disclose *the imaging device as claimed in claim 1, further comprising a light source that is oriented toward the retractor opening so as to illuminate teeth of a patient through the retractor opening* **(Fig. 2A,** reference numerals **216a-d; Fig. 5,** reference numeral **516)**.

Regarding claim **17**, **Prakash et al.** disclose *the imaging device as claimed in claim 16, in which the light source is configured so as to project, through the retractor opening, a reference frame onto the teeth* **(Fig. 5,** reference numeral **516** shows the light source that projects light through the opening **511**. In **[0090], L3-5,** it describes two modes of lighting, e.g. bright field and auto-fluorescent, each of which is equivalent to a reference frame).

Regarding claim **18**, **Prakash et al.** disclose *the imaging device as claimed in claim 17, further comprising a monitoring module configured to monitor properties of radiation emitted by*

the light source as a function of the luminous radiation received by the retractor opening **([0037], L9-16;** it discloses detection of auto-fluorescent radiation after illuminating the teeth area with light of a certain wavelength).

Regarding claim **19**, **Prakash et al.** disclose *the imaging device as claimed in claim 1, in which the image acquisition apparatus has an objective and is positioned with respect to the acquisition opening so that the objective is maintained substantially in the center of the acquisition opening* **([0041]; [0058], L5-9;** it teaches placing a lens (e.g. a fish-eye lens) on the optical path as an objective lens).

Regarding claim **21**, **Prakash et al.** disclose *the imaging device as claimed in claim 1, in which the retractor opening is configured so that both teeth of an upper dental arch of the patient and teeth of a lower arch of the patient are fully visible by the image acquisition apparatus* **(Figs. 9A-C)**.

Regarding claim **22**, **Prakash et al.** disclose *the imaging device as claimed in claim 1, in which the support has a lateral wall which extends between two end faces of the support defining the retractor opening and the acquisition opening, respectively, said lateral wall being rectangular in cross section* (**Fig. 5B** shows the lateral part of the camera mount **520** which is rectangular in cross-section).

Regarding claim **23**, **Prakash et al.** disclose *the imaging device as claimed in claim 1, in which the mechanism for fastening the acquisition apparatus is configured so that the*

acquisition apparatus may be fastened to the support in only one predetermined position **([0062];** it discloses that the adjustable clips **226** enable the image acquisition device to be fastened to the mount **220** in different positions, however, it also discloses that in certain scenario, e.g., for a camera cell phone, the clips are not adjustable along a track, meaning the fastening of the camera is achieved only in one position).

Regarding claim **24**, **Prakash et al.** disclose *the imaging device as claimed in claim 1, in which the retractor includes a rim extending around the retractor opening and arranged in such a way that the patient's lips may rest on it, leaving the patient's teeth visible through said retractor opening* **([0102], L3-10; Fig. 5A** shows the rim along the bite guides **512** which allows the patients lips to rest on it while opening the mouth up).

Regarding claim **25**, **Prakash et al.** disclose *the imaging device as claimed in claim 24, in which the rim has the shape of a channel configured to hold the patient's lips* **([0102], L3-10; Fig. 5A** shows the rim in the shape of a channel along the bite guides **512** which allows the patients lips to rest on it while opening the mouth up).

Regarding claim **26**, **Prakash et al.** disclose *the imaging device as claimed in claim 1, in which the retractor opening is curved around an axis Y which is substantially vertical in a service position* **(Fig. 5A** shows the retractor opening **511,** the four corners of which are curved around an axis perpendicular the plane of the opening).

Regarding claim **27**, **Prakash et al.** disclose *the imaging device as claimed in claim 1, in which the retractor opening is larger than the acquisition opening* **(Fig. 2A** shows the retractor opening **211** larger than the optical path **223** through which the image is acquired by the camera).

Regarding claim **28**, **Prakash et al.** disclose *a patient-operated imaging device* **(Fig. 5B)** *comprising:*

- a support **(Figs. 2A-C,** reference numeral **220. Figs. 5A-B,** reference numeral **520)***;*

- a mouth retractor formed as an integral part of the support and defining a retractor opening **(Figs. 2A-B,** reference numeral **210** is the mouthpiece or retractor and **211** is the opening. **Figs. 5A-B,** reference numeral **510** is the mouthpiece or retractor and **511** is the opening. To be specific, reference numerals **212a-b**, known as bite guides, on the mouthpiece **210** act as the mouth retractors as shown in **Fig. 2A)***; and*

- an image acquisition apparatus **(Fig. 2C,** reference numeral **280)** *fastened to the support in a position in which the image acquisition apparatus is oriented so as to receive an image of the retractor opening* **([0062]; Fig. 2B,** reference numeral **226a-c)**,

wherein the support takes the form of a box **(Fig. 5B,** reference numeral **520** shows the box shaped camera mount as the support**)** *that is in communication with the outside via the retractor opening and via an acquisition opening through which the image acquisition apparatus fastened to the support receives the image of the retractor opening* **([0062]; Fig. 5A** shows the retractor or the mouthpiece **510** and the opening **511** through which the image acquisition device or the cellphone camera receives the image for capturing**)**, *the support being configured so that the image acquisition apparatus observes the retractor opening regardless of the configuration*

of the support **([0069]-[0070]; Figs. 3A-C** show the cellphone camera acquiring the images of the oral cavity through the opening regardless of the configuration of the mount),

wherein the patient-operated imaging device is adapted to obtain a plurality of images **([0069]-[0070]; Figs. 3A-C** show capturing of plurality of images by the cellphone),

wherein at least two of the plurality of images correspond to different angles with respect to the patient's teeth **([0069]-[0070]; Figs. 3A-C** show capturing of plurality of images by the cellphone and as the drawings indicate the plurality of images are taken at different angles with respect to the patient's teeth).

Claim Rejections - 35 USC § 103

The following is a quotation of 35 U.S.C. 103 which forms the basis for all obviousness rejections set forth in this Office action:

> A patent for a claimed invention may not be obtained, notwithstanding that the claimed invention is not identically disclosed as set forth in section 102, if the differences between the claimed invention and the prior art are such that the claimed invention as a whole would have been obvious before the effective filing date of the claimed invention to a person having ordinary skill in the art to which the claimed invention pertains. Patentability shall not be negated by the manner in which the invention was made.

Claim 2 is rejected under 35 U.S.C. 103 as being unpatentable over Prakash et al. (US PGPub 2013/0209954 A1) in view of Charles (US PGPub 2014/0005484 A1).

Regarding claim **2**, **Prakash et al.** teach *the imaging device as claimed in claim 1.*

Although, **Prakash et al.** in **[0054]**, teach that the mouthpiece material is magnetic, and in **[0062]**, it teaches that for a camera cell phone, fastening clips are not adjustable along a track,

meaning the fastening of the camera is achieved only in one position, but it does not explicitly teach fastening mechanism of the image acquisition device is magnetic.

However, **Charles** teach a system in the same field of endeavor **(Figs. 18, 19, 20A)**, where it teaches the fastening mechanism is magnetic **(Charles; [0037], [0263], L13-19)**.

It would have been obvious before the effective filing date of the claimed invention to a person having ordinary skill in the art to combine **Prakash et al**'s invention of techniques of capturing intra-oral images with mobile devices to include **Charles**' usage of magnetic fastening, because this can facilitate convenient attachment within a short period of time **(Charles; [0263], L16-23)**.

Claims 4-11 are rejected under 35 U.S.C. 103 as being unpatentable over Prakash et al. (US PGPub 2013/0209954 A1) in view of Shanjani et al. (US PGPub 2018/0000563 A1).

Regarding claim **4**, **Prakash et al.** teach *the imaging kit as claimed in claim 3.*

But, **Prakash et al.** do not teach that *the imaging device further comprises a detection member and the image acquisition apparatus further comprises a detector that is configured to detect the detection member when the detection member is less than 20 cm from the imaging device.*

However, **Shanjani et al.** teach a system in the same field of endeavor **(Abstract)**, where it teaches *the imaging device further comprises a detection member* **(Shanjani et al.; Fig. 37A** shows a phone having a magnetometer which can detect a magnetic field produced by the NFC device (i.e., detection member is the NFC chip in this example)) *and the image acquisition apparatus further comprises a detector that is configured to detect the detection member when*

the detection member is less than 20 cm from the imaging device **(Shanjani et al.; Fig. 37A** shows a phone having a magnetometer which can detect a magnetic field produced by the NFC device which is less than 20cm from the phone **([0163], [0135], [0278], [0014]**); **[0013]**: The removable mechanical activation interrupt may comprise a magnetic switch, a removable activation rod, a pin, etc. Any of these apparatuses may include the dental appliance (e.g., an aligner such as a shell aligner) to which the monitoring apparatus (e.g., ECI) may be permanently or removably coupled; **[0121]**: The one or more proximity sensors may comprise one or more of: a capacitive sensor, an eddy-current sensor, a magnetic sensor; **[0138]**: the sensors herein can be configured as a switch that is activated and/or deactivated in response to a particular type of signal (e.g., optical, electrical, magnetic, mechanical, etc.); **[0146]**: Alternatively, the intraoral appliance can be operably coupled to a plurality of monitoring devices, such as at least two, three, four, five, or more monitoring devices. Some or all of the monitoring devices may be of the same type (e.g., collect the same type of data). Alternatively, some or all of the monitoring devices may be of different types (e.g., collect different types of data). Any of the embodiments of monitoring devices described herein can be used in combination with other embodiments in a single intraoral appliance).

It would have been obvious before the effective filing date of the claimed invention to a person having ordinary skill in the art to combine **Prakash et al**'s invention of techniques of capturing intra-oral images with mobile devices to include **Shanjani et al**'s detection member, because it creates a device which uses the sensors as a switch that is activated and/or deactivated in response to a particular type of signal (e.g., optical, electrical, magnetic, mechanical, etc.) (**Shanjani et al.; [0138]**).

Regarding claim **5**, **Prakash et al.** and **Shanjani et al.** teach *the imaging kit as claimed in claim 4, in which the detector is configured so as to detect the detection member only when the detection member is less than 20 cm from the imaging device* **(Shanjani et al.; [0163], [0135], [0278], [0014]; Fig. 37A** shows a phone having a magnetometer which can detect a magnetic field produced by the NFC device which is less than 20cm from the phone).

Regarding claim **6**, **Prakash et al.** and **Shanjani et al.** teach *the imaging kit as claimed in claim 4, in which the detection member is positioned less than 5 cm from the edge of the acquisition opening* **(Shanjani et al.; [0163], [0135], [0278], [0014]; Fig. 37A** shows a phone having a magnetometer which can detect a magnetic field produced by the NFC device which is less than 5cm from the phone).

Regarding claim **7**, **Prakash et al.** and **Shanjani et al.** teach *the imaging kit as claimed in claim 4, in which the detector further comprises a magnetometer* **(Shanjani et al.; [0048], [0121], [0138]**).

Regarding claim **8**, **Prakash et al.** and **Shanjani et al.** teach *the imaging kit as claimed in claim 4, in which the detector is configured to trigger an execution of a computer program loaded on a processing module of the image acquisition apparatus in the event that the detection member is detected* **(Shanjani et al.**; **[0145]**: In some embodiments, the monitoring device **300** is dormant before being delivered to the patient (e.g., during storage, shipment, etc.) and is activated only when ready for use. This approach can be beneficial in conserving power expenditure. For example, the components of the monitoring device **300** can be electrically

coupled to the power source **316** at assembly, but may be in a dormant state until activated, e.g., by an external device such as a mobile device, personal computer, laptop, tablet, wearable device, power hub etc. The external device can transmit a signal to the monitoring device **300** that causes the activation mechanism to activate the monitoring device **300**; **[0258]**: As mentioned above, any of the apparatuses described herein (including systems) may communicate with a hand-held electronic device such a smartphone via control software running on the smartphone (or other hand-held electronics). This application software may interface with the electronic compliance indicator and may enhance wireless communications between the electronic compliance indicator (ECI) using NFC and BLE protocols… An ECI apparatus may generally record sensor data from patients wearing an orthodontic appliance such as an aligner. The data may be stored in physical memory on the ECI and retrieved by another device, e.g., using NFC and BLE technologies as described above (or NFC and NFC), so that the smartphone may retrieve the data. The smartphone application (app) may consist of several components, some of which are described in **FIGS. 41, 42** and **43**. For example, in **FIG. 41** schematically illustrates an NFC/BLE communication control. In addition, **FIGS. 44, 45** and **46** schematically illustrate operational states of the ECI device, as well as control of communication between the device and a remote processor (e.g., smartphone)).

It would have been obvious before the effective filing date of the claimed invention to a person having ordinary skill in the art to combine **Prakash et al**'s invention of techniques of capturing intra-oral images with mobile devices to include **Shanjani et al**'s triggering of computer program, because it conserves power (**Shanjani et al.; [0145]**).

Regarding claim **9**, **Prakash et al.** and **Shanjani et al.** teach *the imaging kit as claimed in claim 8, in which the processing module is configured to acquire, in response to the detection of the detection member by the detector, one or more updated images, then analyze the one or more updated images* **(Shanjani et al.; [0142], L8-9)** *to detect an incorrect positioning of the image acquisition apparatus and/or an incorrect positioning of the retractor* **(Shanjani et al.; [0249], L18-26)** *and/or a poor illumination of the retractor opening and/or an unsuitable support length* **(Shanjani et al.; [0145]**: In some embodiments, the monitoring device **300** is dormant before being delivered to the patient (e.g., during storage, shipment, etc.) and is activated only when ready for use. This approach can be beneficial in conserving power expenditure. For example, the components of the monitoring device **300** can be electrically coupled to the power source **316** at assembly, but may be in a dormant state until activated, e.g., by an external device such as a mobile device, personal computer, laptop, tablet, wearable device, power hub etc. The external device can transmit a signal to the monitoring device **300** that causes the activation mechanism to activate the monitoring device **300**; **[0164]**: Some of the proximity sensor types described herein (e.g., capacitive sensors) may also be touch sensors, such that they are activated both by proximity to the sensing target as well as direct contact with the target; **[0171]**: Although **FIG. 8B** illustrates a single monitoring device **850** with a single capacitive sensor **854**, other configurations can also be used. For example, in alternative embodiments, the monitoring device **850** can include multiple capacitive sensors located at different sites on the appliance **852** to detect proximity to and/or contact with multiple locations in the intraoral cavity. Optionally, multiple monitoring devices can be used, with each device being coupled to one or more respective capacitive sensors).

It would have been obvious before the effective filing date of the claimed invention to a person having ordinary skill in the art to combine **Prakash et al**'s invention of techniques of capturing intra-oral images with mobile devices to include **Shanjani et al**'s triggering of computer program, because it conserves power **(Shanjani et al.; [0145])**.

Regarding claim **10**, **Prakash et al.** and **Shanjani et al.** teach *the imaging kit as claimed in claim 4, in which the image acquisition apparatus transmits a message in response to the detection of the detection member by the detector, the message relating to the use of the imaging device and/or relating to the fastening of the acquisition apparatus and/or relating to the fastening of the dental retractor and/or relating to the timing of the updated images to be acquired* **(Shanjani et al.; [0145]**: In some embodiments, the monitoring device **300** is dormant before being delivered to the patient (e.g., during storage, shipment, etc.) and is activated only when ready for use. This approach can be beneficial in conserving power expenditure. For example, the components of the monitoring device **300** can be electrically coupled to the power source **316** at assembly, but may be in a dormant state until activated, e.g., by an external device such as a mobile device, personal computer, laptop, tablet, wearable device, power hub etc. The external device can transmit a signal to the monitoring device **300** that causes the activation mechanism to activate the monitoring device **300)**.

It would have been obvious before the effective filing date of the claimed invention to a person having ordinary skill in the art to combine **Prakash et al**'s invention of techniques of capturing intra-oral images with mobile devices to include **Shanjani et al**'s detection member, because it creates a device which uses the sensors as a switch that is activated and/or deactivated

in response to a particular type of signal (e.g., optical, electrical, magnetic, mechanical, etc.) **(Shanjani et al.; [0138])**.

Regarding claim **11**, **Prakash et al.** teach *the imaging device as claimed in claim 1.*

But it does not explicitly teach that *the mechanism for fastening includes the detection member.*

However, **Shanjani et al.** teach a system in the same field of endeavor **(Abstract)**, where it teaches that *the mechanism for fastening includes the detection member* **(Shanjani et al.; [0148])**.

It would have been obvious before the effective filing date of the claimed invention to a person having ordinary skill in the art to combine **Prakash et al**'s invention of techniques of capturing intra-oral images with mobile devices to include **Shanjani et al**'s usage of the detection member on the fastener, because it is beneficial to distribute the components of the monitoring device across multiple appliances in order to accommodate space limitations, accommodate power limitations, and/or improve sensing **(Shanjani et al.; [0152], L12-16)**.

Claim 20 is rejected under 35 U.S.C. 103 as being unpatentable over Prakash et al. (US PGPub 2013/0209954 A1) in view of Pfeiffer (US Pat 5,677,537).

Regarding claim **20**, **Prakash et al.** teach *the imaging device as claimed in claim 1.*

But **Prakash et al.** do not explicitly teach the mechanism of an elastic member.

However, **Pfeiffer** teaches a system in the same field of endeavor **(Abstract)**, where it teaches the fastening mechanism is made of an elastic member **(Pfeiffer; Fig. 2, Col 3, L40-45)**.

It would have been obvious before the effective filing date of the claimed invention to a person having ordinary skill in the art to combine **Prakash et al**'s invention of techniques of capturing intra-oral images with mobile devices to include **Pfeiffer**'s usage of elastic material, because the holder (retractor) and sensor (camera) are secured in a defined alignment relative to one another **(Pfeiffer; Fig. 2, Col 4, L4-8)**.

Conclusion

The prior art made of record and not relied upon is considered pertinent to applicant's disclosure.

1. "CHEEK RETRACTOR AND MOBILE DEVICE HOLDER" – Meyer et al., US PGPub 2018/0228359 A1.

2. "IMAGING DEVICE FOR DENTAL INSTRUMENTS AND METHODS FOR INTRA-ORAL VIEWING" – Karazivan et al., US PGPub 2012/0040305 A1.

3. "INTRA-ORAL CAMERA" – Matthews, US Pat 9939714 B1.

4. "METHODS AND APPARATUSES FOR DENTAL IMAGES" – Carrier, Jr. et al., US PGPub 2018/0125610 A1.

5. "SOFT HEAD MOUNTED DISPLAY GOGGLES FOR USE WITH MOBILE COMPUTING DEVICES" – Lyons, US PGPub 2015/0234192 A1.

6. "Dental Informatics and Intra-oral Photography in Communicating with Dental Students in the Dominican Republic" - Lawrence PARRISH, Anton DIY, Nicholas R. KENNING, Kristen TEMPLETON, Ruben SAGUN, Nicole S. KIMMES, Gene GASPARD, Stephen J. HESS; Journal of Health Informatics in Developing Countries; April 30, 2014.

7. "A 3-D Reconstruction System for the Human Jaw Using a Sequence of Optical Images" - Sameh M. Yamany, Aly A. Farag, David Tasman, Allan G. Farman; IEEE TRANSACTIONS ON MEDICAL IMAGING, VOL. 19, NO. 5, MAY 2000.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to MAINUL HASAN whose telephone number is (571)272-0422. The examiner can normally be reached on MON-FRI: 10AM-6PM, Alternate FRIDAYS, EST.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, JAY PATEL can be reached on (571)272-2988. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/Mainul Hasan/
Primary Examiner, Art Unit 2485

EXHIBIT R-10

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
16/921,545	07/06/2020	Philippe SALAH	N&P-51400US1	8521

108676 7590 01/12/2022

Ronald M. Kachmarik
Cooper Legal Group LLC
1388 Ridge Road, Unit 1
Hinckley, OH 44233

EXAMINER	
HASAN, MAINUL	

ART UNIT	PAPER NUMBER
2485	

NOTIFICATION DATE	DELIVERY MODE
01/12/2022	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

docketing@cooperlegalgroup.com

Office Action Summary	Application No. 16/921,545	Applicant(s) SALAH et al.	
	Examiner MAINUL HASAN	Art Unit 2485	AIA (FITF) Status Yes

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTHS FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☑ Responsive to communication(s) filed on 30 December 2021.

☐ A declaration(s)/affidavit(s) under **37 CFR 1.130(b)** was/were filed on _____.

2a)☐ This action is **FINAL.** 2b) ☑ This action is non-final.

3)☐ An election was made by the applicant in response to a restriction requirement set forth during the interview on _____; the restriction requirement and election have been incorporated into this action.

4)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims*

5) ☑ Claim(s) 1-20 is/are pending in the application.

5a) Of the above claim(s) _____ is/are withdrawn from consideration.

6) ☐ Claim(s) _____ is/are allowed.

7) ☑ Claim(s) 1-20 is/are rejected.

8) ☐ Claim(s) _____ is/are objected to.

9) ☐ Claim(s) _____ are subject to restriction and/or election requirement

* If any claims have been determined allowable, you may be eligible to benefit from the **Patent Prosecution Highway** program at a participating intellectual property office for the corresponding application. For more information, please see http://www.uspto.gov/patents/init_events/pph/index.jsp or send an inquiry to **PPHfeedback@uspto.gov.**

Application Papers

10)☑ The specification is objected to by the Examiner.

11)☑ The drawing(s) filed on 16 March 2021 is/are: a)☑ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

Priority under 35 U.S.C. § 119

12)☑ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

Certified copies:

a)☑ All b)☐ Some** c)☐ None of the:

1.☑ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

** See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☑ Notice of References Cited (PTO-892)

2) ☑ Information Disclosure Statement(s) (PTO/SB/08a and/or PTO/SB/08b) Paper No(s)/Mail Date _____.

3) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date _____.

4) ☐ Other: _____.

DETAILED ACTION

Notice of Pre-AIA or AIA Status

The present application, filed on or after March 16, 2013, is being examined under the first inventor to file provisions of the AIA. In the event the determination of the status of the application as subject to AIA 35 U.S.C. 102 and 103 (or as subject to pre-AIA 35 U.S.C. 102 and 103) is incorrect, any correction of the statutory basis for the rejection will not be considered a new ground of rejection if the prior art relied upon, and the rationale supporting the rejection, would be the same under either status. There are a total of 20 claims and claims 1-20 are pending.

Examiner's Response to Preliminary Amendments

The Examiner acknowledges and enters for consideration the Preliminary Claim Amendments filed on 07/06/2020 and Preliminary Drawing Amendments filed on 03/16/2021. Claims 21-64 have been cancelled. Claims 1-20 remain pending in the current application.

Priority

Acknowledgment is made of applicant's claim for foreign priority based on an application filed in FR1753392 on 04/19/2017, FR1753389 04/19/2017, and EP17306361.1 10/10/2017. Receipt is acknowledged of certified copies of papers required by 37 CFR 1.55.

Claim Objections

Claim 7 is objected to because of the following informalities:

Claim 7 recites "*…which extends between two end faces defining said first opening and said second opening* ***to***". It is not clear what the word "to" is contributing to the limitation. The Examiner believes the word should be removed.

Appropriate correction is required.

Claim Rejections - 35 USC § 112

The following is a quotation of 35 U.S.C. 112(b):

(B) CONCLUSION.—The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the inventor or a joint inventor regards as the invention.

The following is a quotation of 35 U.S.C. 112, second paragraph:

The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.

Claims 8, 20 are rejected under 35 U.S.C. 112(b) or 35 U.S.C. 112 (pre-AIA), second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which the inventor or a joint inventor (or for applications subject to pre-AIA 35 U.S.C. 112, the applicant), regards as the invention.

The term "substantially" in claims 8, 20 is a relative term which renders the claim indefinite. The term "substantially" is not defined by the claim, the specification does not provide a standard for ascertaining the requisite degree, and one of ordinary skill in the art would not be reasonably apprised of the scope of the invention.

Double Patenting

The nonstatutory double patenting rejection is based on a judicially created doctrine grounded in public policy (a policy reflected in the statute) so as to prevent the unjustified or improper timewise extension of the "right to exclude" granted by a patent and to prevent possible harassment by multiple assignees. A nonstatutory double patenting rejection is appropriate where the conflicting claims are not identical, but at least one examined application claim is not patentably distinct from the reference claim(s) because the examined application claim is either anticipated by, or would have been obvious over, the reference claim(s). See, e.g., *In re Berg*, 140 F.3d 1428, 46 USPQ2d 1226 (Fed. Cir. 1998); *In re Goodman*, 11 F.3d 1046, 29 USPQ2d 2010 (Fed. Cir. 1993); *In re Longi*, 759 F.2d 887, 225 USPQ 645 (Fed. Cir. 1985); *In re Van Ornum*, 686 F.2d 937, 214 USPQ 761 (CCPA 1982); *In re Vogel*, 422 F.2d 438, 164 USPQ 619 (CCPA 1970); *In re Thorington*, 418 F.2d 528, 163 USPQ 644 (CCPA 1969).

A timely filed terminal disclaimer in compliance with 37 CFR 1.321(c) or 1.321(d) may be used to overcome an actual or provisional rejection based on nonstatutory double patenting provided the reference application or patent either is shown to be commonly owned with the examined application, or claims an invention made as a result of activities undertaken within the scope of a joint research agreement. See MPEP § 717.02 for applications subject to examination under the first inventor to file provisions of the AIA as explained in MPEP § 2159. See MPEP § 2146 *et seq.* for applications not subject to examination under the first inventor to file provisions of the AIA. A terminal disclaimer must be signed in compliance with 37 CFR 1.321(b).

The USPTO Internet website contains terminal disclaimer forms which may be used. Please visit www.uspto.gov/patent/patents-forms. The filing date of the application in which the form is filed determines what form (e.g., PTO/SB/25, PTO/SB/26, PTO/AIA/25, or

PTO/AIA/26) should be used. A web-based eTerminal Disclaimer may be filled out completely online using web-screens. An eTerminal Disclaimer that meets all requirements is auto-processed and approved immediately upon submission. For more information about eTerminal Disclaimers, refer to www.uspto.gov/patents/process/file/efs/guidance/eTD-info-I.jsp.

Claims **1-12** are rejected on the ground of nonstatutory double patenting as being unpatentable over claims 1-17 of U.S. Patent No. **10,736,715 B2**. Although the claims at issue are not identical, they are not patentably distinct from each other because of the following reasons.

Claim **1** of the instant application is rejected on the ground of nonstatutory double patenting as being unpatentable over claim 1 of U.S. Patent No. **10,736,715 B2**. Although the claims at issue are not identical, they are not patentably distinct from each other because of the following reasons:

	16921545 (Instant Application)	**10,736,715 B2** (Patent)
	Claim 1	**Claim 1**
1	*A method to* ***acquire dental images of a patient with a support defining a chamber that is in communication with an outside of said chamber via a first opening and via a second opening****, said method comprising the following steps:*	*An imaging device including:*
2	*-* ***fixing a mobile phone in front of the second opening;***	***a support;***
3	*-* ***positioning said first opening in front of a mouth of the patient;***	***a mouth retractor fastened to the support and defining a retractor opening; and***
4	*-* ***acquiring at least one dental image by means of the mobile phone****.*	***a mechanism for fastening an image acquisition apparatus to the support in a position in which the acquisition apparatus is oriented so as to receive an image of the retractor opening,***

5		***wherein the support takes the form of a box that is in communication with the outside via the retractor opening and via an acquisition opening through which the acquisition apparatus fastened to the support receives said image of the retractor opening,***
6		*the support being configured so that the acquisition apparatus observes the retractor opening regardless of a configuration of said support,*
7		*said mechanism being chosen from the group consisting of clip-fastening means, self-gripping strips of hook and loop fastener type, clamping jaws, screws, magnets, and complementarity of shape between the support and the acquisition apparatus, or consisting of a cover that may be clamped against the support.*

The equivalence in claim limitations of the instant application and the patent are highlighted in **bold** texts. Although the instant application claim limitations appear to be a broader version of the corresponding Patent claim limitation, however, a close review shows that the limitations are identical. Although the instant application claims a method and the Patent claims a device, however, they are not patentable different from each other because of the following explanation: The method describes dental image acquisition of a patient by a support defining a chamber having a first end a second end for communicating with the outside. The Patent describes an imaging device having a support in the shape of a box (chamber of instant application) that communicates with the outside through a retractor opening (first opening of instant application) and an acquisition opening (second opening of instant application). The method also describes fixing a mobile phone in front of the second opening. The Patent describes an image acquisition apparatus (mobile phone of instant application since the mobile phone is

exclusively used for acquiring an image) which is placed at the acquisition opening through which the acquisition apparatus fastened to the support receives said image. The method then describes positioning the first opening in front of a mouth of the patient. The Patent describes a mouth retractor (which goes in to retract the patient's mouth) fastened to the support and defining a retractor opening (first opening of the instant application). Lastly, the method describes acquiring a dental image by means of the mobile phone. The Patent describes the acquisition apparatus (mobile phone) fastened to the support receives said image of the retractor opening (acquires dental image of the mouth). Therefore, the instant application claim 1 as a whole is not patentable over the Patent claim 1. This is a non-statutory double patenting rejection.

Claims **2-12** of the instant application are rejected on the ground of nonstatutory double patenting as being unpatentable over a combination of claims of U.S. Patent No. **10,736,715 B2**.

Claims **13-20** are rejected on the ground of nonstatutory double patenting as being unpatentable over claims 1-17 of U.S. Patent No. **10,736,715 B2** in view of **Dorodvand et al. (US PGPub 2019/0167115 A1) (Disclosed in IDS)**. Although the claims at issue are not identical, the instant application claim is not patentable over the Patent claims in view of **Dorodvand et al**.

Claim **13** of the instant application is rejected on the ground of nonstatutory obviousness type double patenting as being unpatentable over claim 1 of U.S. Patent No. **10,736,715 B2** in view of **Dorodvand et al. (US PGPub 2019/0167115 A1) (Disclosed in IDS)**. Although the

claims at issue are not identical, the instant application claim is not patentable over the Patent claim in view of **Dorodvand et al.** as shown in the following table:

	16921545 (Instant Application)	**10,736,715 B2** (Patent)
	Claim 13	**Claim 1**
1	*a method to* ***acquire dental images of a patient with a support defining a chamber that is in communication with an outside of said chamber via a first opening and via a second opening,*** *the distance between said openings being constant, said method comprising the following steps:*	***An imaging device including:***
2	***- fixing a mobile phone in front of the second opening, in one predetermined position,***	***a support;***
3	***- positioning a mouth of the patient in front of the first opening;***	***a mouth retractor fastened to the support and defining a retractor opening; and***
4	***- acquiring, by the patient, at least one dental image by means of the mobile phone.***	***a mechanism for fastening an image acquisition apparatus to the support in a position in which the acquisition apparatus is oriented so as to receive an image of the retractor opening,***
5		***wherein the support takes the form of a box that is in communication with the outside via the retractor opening and via an acquisition opening through which the acquisition apparatus fastened to the support receives said image of the retractor opening,***
6		*the support being configured so that the acquisition apparatus observes the retractor opening regardless of a configuration of said support,*

7		*said mechanism being chosen from the group consisting of clip-fastening means, self-gripping strips of hook and loop fastener type, clamping jaws, screws, magnets, and complementarity of shape between the support and the acquisition apparatus, or consisting of a cover that may be clamped against the support.*

The equivalence in claim limitations of the instant application and the patent are highlighted in **bold** texts. Although the instant application claim limitations appear to be a broader version of the corresponding Patent claim limitation, however, a close review shows that the limitations are identical except the limitation of the distance between the first opening and second opening being constant, which is not present in the Patent. The rest of the instant application limitations are identical to the Patent limitations (Please see the explanation for claim 1 DP rejection above). However, **Dorodvand et al.** teach a system in the same field of endeavor **(Abstract)**, where it teaches the limitation of the distance between the first opening and second opening being constant **(Dorodvand et al.; [0042], L1-8)**. It would have been obvious before the effective filing date of the claimed invention to a person having ordinary skill in the art to combine the Patent's invention of a dental imaging device to include **Dorodvand et al**'s usage of constant distance between the image acquisition device and the mouth, because the image area is therefore constant between images. Similarly the distance of the image capture device from the teeth and gums is constant providing a consistent focal length and rotation between images **(Dorodvand et al.; [0042], L1-8)**. Therefore, the instant application claim 13 as a whole is not patentable over the patent claim 1 in view of **Dorodvand et al.** This is a non-statutory obviousness type double patenting rejection.

Claims **14-20** of the instant application are rejected on the ground of nonstatutory obviousness type double patenting as being unpatentable over a combination of claims of U.S. Patent No. **10,736,715 B2** in view of **Dorodvand et al. (US PGPub 2019/0167115 A1) (Disclosed in IDS)**.

Claims **1-12** of the instant application are also rejected on the ground of nonstatutory double patenting as being unpatentable over claims 1-21 of U.S. Patent No. **10,842,592 B2**.

Claim **1** of the instant application is rejected on the ground of nonstatutory double patenting as being unpatentable over claim 1 of U.S. Patent No. **10,842,592 B2**. Although the claims at issue are not identical, they are not patentably distinct from each other because of the following reasons:

	16921545 (Instant Application)	**10,842,592 B2** (Patent)
	Claim 1	**Claim 1**
1	*A method to* ***acquire dental images of a patient with a support defining a chamber that is in communication with an outside of said chamber via a first opening and via a second opening****, said method comprising the following steps:*	***An imaging device including:***
2	*-* ***fixing a mobile phone in front of the second opening;***	***a support;***
3	*-* ***positioning said first opening in front of a mouth of the patient;***	***a mouth retractor fastened, to the support and defining a retractor opening****, the mouth retractor including a rim extending around the retractor opening; a colorimetric calibration chart and/or a translucence calibration chart; and a light source that is oriented so as to illuminate both teeth of a patient through the retractor opening and said colorimetric calibration chart and/or said translucence calibration chart;*

4	- ***acquiring at least one dental image by means of the mobile phone***.	***means for fastening an image acquisition apparatus to the support in a position in which the image acquisition apparatus is oriented so as to receive an image of*** *both* ***the retractor opening*** *and of said colorimetric calibration chart and/or of said translucence calibration chart,*
5		*wherein said means for fastening can be deactivated, the mouth retractor being configured so as, in a service position in which the* ***mouth retractor is positioned on the mouth of a patient****, lips of the patient may rest on said rim so that only an inside of the mouth is visible through the retractor opening;*
6		***the support taking the form of a box, that is in communication with the outside via the retractor opening and via an acquisition opening through which the acquisition apparatus fastened to the support receives said image of the retractor opening,***
7		***the support being configured so that the acquisition apparatus observes the retractor opening regardless of a configuration of said support; the image acquisition apparatus being a mobile phone or a tablet.***

The equivalence in claim limitations of the instant application and the patent are highlighted in **bold** texts. Although the instant application claim limitations appear to be a broader version of the corresponding Patent claim limitation, however, a close review shows that the limitations are identical. Although the instant application claims a method and the Patent claims a device, however, they are not patentable different from each other because of the following explanation: The method describes dental image acquisition of a patient by a support

defining a chamber having a first end a second end for communicating with the outside. The Patent describes an imaging device having a support in the shape of a box (chamber of instant application) that communicates with the outside through a retractor opening (first opening of instant application) and an acquisition opening (second opening of instant application). The method also describes fixing a mobile phone in front of the second opening. The Patent describes the image acquisition apparatus being a mobile phone or a tablet which is placed at the acquisition opening (second opening of instant application) through which the acquisition apparatus fastened to the support receives said image. The method then describes positioning the first opening in front of a mouth of the patient. The Patent describes a mouth retractor (which goes in to retract the patient's mouth) wherein the mouth retractor is positioned on the mouth of a patient, and fastened to the support and defining a retractor opening (first opening of the instant application). Lastly, the method describes acquiring a dental image by means of the mobile phone. The Patent describes the acquisition apparatus (mobile phone) fastened to the support receives said image of the retractor opening (acquires dental image of the mouth). Therefore, the instant application claim 1 as a whole is not patentable over the Patent claim 1. This is a non-statutory double patenting rejection.

Claims **2-12** of the instant application are also rejected on the ground of nonstatutory double patenting as being unpatentable over a combination of claims of U.S. Patent No. **10, 842,592 B2**.

Claim **13** of the instant application is rejected on the ground of nonstatutory obviousness type double patenting as being unpatentable over claim 1 of U.S. Patent No. **10,842,592 B2** in view of **Dorodvand et al. (US PGPub 2019/0167115 A1) (Disclosed in IDS)**. Although the

claims at issue are not identical, the instant application claim is not patentable over the Patent claim in view of **Dorodvand et al.** as shown in the following table:

	16921545 (Instant Application)	**10,842,592 B2** (Patent)
	Claim 13	**Claim 1**
1	*a method to* ***acquire dental images of a patient with a support defining a chamber that is in communication with an outside of said chamber via a first opening and via a second opening,*** *the distance between said openings being constant, said method comprising the following steps:*	***An imaging device including:***
2	***- fixing a mobile phone in front of the second opening, in one predetermined position,***	***a support;***
3	***- positioning a mouth of the patient in front of the first opening;***	***a mouth retractor fastened, to the support and defining a retractor opening****, the mouth retractor including a rim extending around the retractor opening; a colorimetric calibration chart and/or a translucence calibration chart; and a light source that is oriented so as to illuminate both teeth of a patient through the retractor opening and said colorimetric calibration chart and/or said translucence calibration chart;*
4	***- acquiring, by the patient, at least one dental image by means of the mobile phone.***	***means for fastening an image acquisition apparatus to the support in a position in which the image acquisition apparatus is oriented so as to receive an image of*** *both* ***the retractor opening*** *and of said colorimetric calibration chart and/or of said translucence calibration chart,*
5		*wherein said means for fastening can be deactivated, the mouth retractor being configured so as, in a service position in which the* ***mouth retractor is positioned on the mouth of a patient****, lips of the patient may rest on said rim so that only an inside*

		of the mouth is visible through the retractor opening;
6		***the support taking the form of a box, that is in communication with the outside via the retractor opening and via an acquisition opening through which the acquisition apparatus fastened to the support receives said image of the retractor opening,***
7		***the support being configured so that the acquisition apparatus observes the retractor opening regardless of a configuration of said support; the image acquisition apparatus being a mobile phone or a tablet.***

The equivalence in claim limitations of the instant application and the patent are highlighted in **bold** texts. Although the instant application claim limitations appear to be a broader version of the corresponding Patent claim limitation, however, a close review shows that the limitations are identical except the limitation of the distance between the first opening and second opening being constant, which is not present in the Patent. The rest of the instant application limitations are identical to the Patent limitations (Please see the explanation for claim 1 DP rejection above). However, **Dorodvand et al.** teach a system in the same field of endeavor **(Abstract)**, where it teaches the limitation of the distance between the first opening and second opening being constant **(Dorodvand et al.; [0042], L1-8)**. It would have been obvious before the effective filing date of the claimed invention to a person having ordinary skill in the art to combine the Patent's invention of a dental imaging device to include **Dorodvand et al**'s usage of constant distance between the image acquisition device and the mouth, because the image area is

therefore constant between images. Similarly the distance of the image capture device from the teeth and gums is constant providing a consistent focal length and rotation between images **(Dorodvand et al.; [0042], L1-8)**. Therefore, the instant application claim 13 as a whole is not patentable over the patent claim 1 in view of **Dorodvand et al.** This is a non-statutory obviousness type double patenting rejection.

Claims **14-20** of the instant application are also rejected on the ground of nonstatutory obviousness type double patenting as being unpatentable over a combination of claims of U.S. Patent No. **10, 842,592 B2** in view of **Dorodvand et al**.

Claim Rejections - 35 USC § 102

The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

> A person shall be entitled to a patent unless –
>
> (a)(1) the claimed invention was patented, described in a printed publication, or in public use, on sale or otherwise available to the public before the effective filing date of the claimed invention.

Claims 1-6, 9-10 are rejected under AIA 35 U.S.C. 102(a)(1) as being anticipated by Prakash et al. (US PGPub 2013/0209954 A1).

Regarding claim **1** (Original), **Prakash et al.** disclose *a method to acquire dental images of a patient* **(Figs. 9A-F)** *with a support* **(Figs. 2A-C,** reference numeral **220. Figs. 5A-B,** reference numeral **520)** *defining a chamber* **(Fig. 5B,** reference numeral **520** shows the chamber which is a box shaped camera mount as the support) *that is in communication with an outside of*

said chamber via a first opening and via a second opening **([0104], L5-10; Fig. 5C** shows the retractor or the mouthpiece **560** and a first opening **571** through which the image acquisition device or the cellphone camera receives the image for capturing and a second opening **561** which opens to the mouth)*, said method comprising the following steps:*

- *fixing a mobile phone in front of the second opening* **(Fig. 5B,** reference numeral **580; Fig. 5C,** reference numeral **584)***;*

- *positioning said first opening in front of a mouth of the patient* **(Fig. 5B** shows the positioning of the mouthpiece's (**510**) other opening in front of the mouth. **Fig. 5C** also shows the first opening **571** through which the cellphone is mounted and second opening **561** which is positioned in front of the mouth)*;*

- *acquiring at least one dental image by means of the mobile phone* **([0040]; Figs. 9A-F**).

Regarding claim **2** (Original), **Prakash et al.** disclose *the method as claimed in claim 1, comprising a step before acquiring at least one dental image, in which a dental retractor is introduced in the mouth of the patient* **(Fig. 5B,** reference numeral **510, Fig. 5C-D,** reference numeral **560** show the mouthpiece which acts as a mouth retractor before acquiring any images).

Regarding claim **3** (Original), **Prakash et al.** disclose *the method as claimed in claim 2, in which the dental retractor is fixed on the support in front of the first opening* (**Fig. 5C** shows the retractor **560** is fixed in front of the first opening **571**).

Regarding claim **4** (Original), **Prakash et al.** disclose *the method as claimed in claim 2, in which the dental retractor is formed as an integral part of the support* (**Figs. 2A-B,** reference numeral **210** is the mouthpiece or retractor and **211** is the opening. **Figs. 5A-B,** reference numeral **510** is the mouthpiece or retractor and **511** is the opening. To be specific, reference numerals **212a-b**, known as bite guides, on the mouthpiece **210** act as the mouth retractors as shown in **Fig. 2A**).

Regarding claim **5** (Original), **Prakash et al.** disclose *the method as claimed in claim 1, in which said positioning comprises a positioning of patient's lips on a rim extending around the first opening* (**Fig. 2A,** reference numerals **212a-b** and **Fig. 5A,** reference numerals **512** represent the bite guides which is equivalent to a rim that act as a spreader of the patient's lips as described in **[0004], L3-10.** Also, see **Fig. 10B,** reference numeral **1012**).

Regarding claim **6** (Original), **Prakash et al.** disclose *the method as claimed in claim 1, comprising an automatic guidance of a user to help positioning of the mouth relative to the support, and/or specifying a number of images to be acquired* (**[0045], L9-12**).

Regarding claim **9** (Original), **Prakash et al.** disclose *the method as claimed in claim 1, in which the support is rectangular in cross section* (**Fig. 5B** shows the camera mount **520** which is rectangular in cross-section).

Regarding claim **10** (Original), **Prakash et al.** disclose *the method as claimed in claim 1, in which the fixing of the mobile phone on the support is performed by fastening the mobile*

phone to the support in only one predetermined position **([0062];** it discloses that the adjustable clips **226** enable the image acquisition device to be fastened to the mount **220** in different positions, however, it also discloses that in certain scenario, e.g., for a camera cell phone, the clips are not adjustable along a track, meaning the fastening of the camera is achieved only in one position).

Claim Rejections - 35 USC § 103

The following is a quotation of 35 U.S.C. 103 which forms the basis for all obviousness rejections set forth in this Office action:

> A patent for a claimed invention may not be obtained, notwithstanding that the claimed invention is not identically disclosed as set forth in section 102, if the differences between the claimed invention and the prior art are such that the claimed invention as a whole would have been obvious before the effective filing date of the claimed invention to a person having ordinary skill in the art to which the claimed invention pertains. Patentability shall not be negated by the manner in which the invention was made.

Claims 7, 11-20 are rejected under 35 U.S.C. 103 as being unpatentable over Prakash et al. (US PGPub 2013/0209954 A1) in view of Dorodvand et al. (US PGPub 2019/0167115 A1) (Disclosed in IDS).

Regarding claim **7** (Original), **Prakash et al.** teach *the method as claimed in claim 1, in which the support comprises lateral wall which extends between two end faces defining said first opening and said second opening to* (**Fig. 10A** shows the optical path opening **1023** created by the support wall (on two sides) which is laterally extended between the two openings).

Although, **Prakash et al.** teach determining distance from an imaging plane to a surface of an oral cavity of a subject as described in **[0010]**, but it does not explicitly teach that the distance between the two openings being constant.

However, **Dorodvand et al.** teach a system in the same field of endeavor **(Abstract)**, where it teaches the distance between the first opening and second opening being constant **(Dorodvand et al.; [0042], L1-8)**.

It would have been obvious before the effective filing date of the claimed invention to a person having ordinary skill in the art to combine **Prakash et al**'s invention of a dental imaging device to include **Dorodvand et al**'s usage of constant distance between the image acquisition device and the mouth, because the image area is therefore constant between images and similarly the distance of the image capture device from the teeth and gums is constant providing a consistent focal length and rotation between images **(Dorodvand et al.; [0042], L1-8)**.

Regarding claim **11** (Original), **Prakash et al.** teach *the method as claimed in claim 1.*

Although, **Prakash et al.** teach acquiring images of a patient's oral cavity using a mobile phone camera mounted on a support attached to the mouth of the patient, but it does not explicitly teach *at least said acquiring of said at least one dental image is performed by the patient.*

However, **Dorodvand et al.** teach a system in the same field of endeavor **(Abstract)**, where it teaches acquisition of dental images being performed by the patient **(Dorodvand et al.; [0021], L16-22)**

It would have been obvious before the effective filing date of the claimed invention to a person having ordinary skill in the art to combine **Prakash et al**'s invention of a dental imaging

device to include **Dorodvand et al**'s capability of patient performing the image acquisition, because the user may reliably and reproducibly capture images of their mouth, without the aid of a skilled operator **(Dorodvand et al.; [0020])**.

Regarding claim **12** (Original), **Prakash et al.** teach *the method as claimed in claim 1.*

But **Prakash et al.** do not explicitly teach *acquiring is performed in less than a minute, without recourse to a specialist.*

However, **Dorodvand et al.** teach a system in the same field of endeavor **(Abstract)**, where it teaches *acquiring is performed in less than a minute, without recourse to a specialist* **(Dorodvand et al.; [0017];** It teaches that the invention seeks to give instant feedback to a user regarding their oral health, which is less than a minute. On the other hand, it also teaches that the invention seeks to provide a straightforward method which can be implemented by an untrained user to provide a quick and accurate estimate of tooth).

It would have been obvious before the effective filing date of the claimed invention to a person having ordinary skill in the art to combine **Prakash et al**'s invention of a dental imaging device to include **Dorodvand et al**'s capability of patient performing the image acquisition, because the user may reliably and reproducibly capture images of their mouth, without the aid of a skilled operator **(Dorodvand et al.; [0020])**.

Regarding claim **13** (Original), **Prakash et al.** teach *a method to acquire dental images of a patient* **(Figs. 9A-F)** *with a support* **(Figs. 2A-C,** reference numeral **220. Figs. 5A-B,** reference numeral **520)** *defining a chamber* **(Fig. 5B,** reference numeral **520** shows the chamber which is a box shaped camera mount as the support) *that is in communication with an outside of said*

chamber via a first opening and via a second opening **([0104], L5-10; Fig. 5C** shows the retractor or the mouthpiece **560** and a first opening **571** through which the image acquisition device or the cellphone camera receives the image for capturing and a second opening **561** which opens to the mouth)*, the distance between said openings being constant, said method comprising the following steps:*

- fixing a mobile phone in front of the second opening, in one predetermined position **(Fig. 5B,** reference numeral **580; Fig. 5C,** reference numeral **584**. The predetermined position could be any of the three positions as shown in **Figs. 3A-C**),

- positioning a mouth of the patient in front of the first opening **(Fig. 5B** shows the positioning of the mouthpiece's (**510**) other opening in front of the mouth. **Fig. 5C** also shows the first opening **571** through which the cellphone is mounted and second opening **561** which is positioned in front of the mouth)*;*

- acquiring, by the patient, at least one dental image by means of the mobile phone **([0040]; Figs. 9A-F;** In **[0108],** it teaches usage of the system by user/clinician. Here the user is analogous to a patient).

Although, **Prakash et al.** teach determining distance from an imaging plane to a surface of an oral cavity of a subject as described in **[0010]**, but it does not explicitly teach that the distance between the two openings being constant.

However, **Dorodvand et al.** teach a system in the same field of endeavor **(Abstract)**, where it teaches the distance between the first opening and second opening being constant **(Dorodvand et al.; [0042], L1-8)**.

It would have been obvious before the effective filing date of the claimed invention to a person having ordinary skill in the art to combine **Prakash et al**'s invention of a dental imaging

device to include **Dorodvand et al**'s usage of constant distance between the image acquisition device and the mouth, because the image area is therefore constant between images and similarly the distance of the image capture device from the teeth and gums is constant providing a consistent focal length and rotation between images (**Dorodvand et al.; [0042], L1-8**).

Regarding claim **14** (Original), **Prakash et al.** and **Dorodvand et al.** teach *the method as claimed in claim 13, in which said positioning comprises positioning of the patient's lip on a rim of the support* (**Prakash et al.; Fig. 2A,** reference numerals **212a-b** and **Fig. 5A,** reference numerals **512** represent the bite guides which is equivalent to a rim that act as a spreader of the patient's lips as described in **[0004], L3-10.** Also, see **Fig. 10B,** reference numeral **1012**).

Regarding claim **15** (Original), **Prakash et al.** and **Dorodvand et al.** teach *the method as claimed in claim 13, comprising a step before acquiring at least one dental image, in which a dental retractor is introduced in the mouth of the patient* (**Prakash et al.; Fig. 5B,** reference numeral **510, Fig. 5C-D,** reference numeral **560** show the mouthpiece which acts as a mouth retractor before acquiring any images).

Regarding claim **16** (Original), **Prakash et al.** and **Dorodvand et al.** teach *the method as claimed in claim 15, in which the dental retractor is fixed on the support in front of the first opening* (**Prakash et al.; Fig. 5C** shows the retractor **560** is fixed in front of the first opening **571**).

Regarding claim **17** (Original), **Prakash et al.** and **Dorodvand et al.** teach *the method as claimed in claim 15, in which the dental retractor is formed as an integral part of the support* **(Prakash et al.; Figs. 2A-B,** reference numeral **210** is the mouthpiece or retractor and **211** is the opening. **Figs. 5A-B,** reference numeral **510** is the mouthpiece or retractor and **511** is the opening. To be specific, reference numerals **212a-b**, known as bite guides, on the mouthpiece **210** act as the mouth retractors as shown in **Fig. 2A**).

Regarding claim **18** (Original), **Prakash et al.** and **Dorodvand et al.** teach *the method as claimed in claim 13, in which the patient is automatically guided for said positioning and/or is specified of a number of images to be acquired* **(Prakash et al.; [0045], L9-12)**.

Regarding claim **19** (Original), **Prakash et al.** and **Dorodvand et al.** teach *the method as claimed in claim 13, in which the support comprises lateral wall which extends between two end faces* **(Prakash et al.; Fig. 10A** shows the optical path opening **1023** created by the support wall (on two sides) which is laterally extended between the two openings)

Regarding claim **20** (Original), **Prakash et al.** and **Dorodvand et al.** teach *the method as claimed in claim 19, in which the lateral wall is substantially cylindrical* **(Prakash et al.; Fig. 10B)**.

Claim 8 is rejected under 35 U.S.C. 103 as being unpatentable over Prakash et al. (US PGPub 2013/0209954 A1) in view of Charles (US PGPub 2014/0005484 A1).

Regarding claim **8** (Original), **Prakash et al.** teach *the method as claimed in claim 1.*

Although **Prakash et al.** show a rectangular cross-section of the chamber as shown in **Fig. 2B-C, 5B-D, 10A,** but it does not explicitly teach that the *chamber is substantially cylindrical.*

However, **Charles**, in the same field of endeavor **(Abstract)**, teaches *chamber is substantially cylindrical* **(Charles; [0328], L6-9)**.

Before the effective filing date of the claimed invention, it would have been a matter of design choice to a person of ordinary skill in the art to use a cylindrical cross-section of the chamber because Applicant has not disclosed that using a cylindrical cross-section chamber provides an advantage, is used for a particular purpose, or solves a stated problem. One of ordinary skill in the art, furthermore, would have expected Applicant's invention to perform equally well with using a cylindrical cross-section chamber because of mere design choice. Therefore, it would have been a design choice to modify **Prakash et al**'s invention of a dental imaging device to include a dental imaging device as taught by **Charles** to obtain the invention as specified in the claim(s).

Conclusion

The prior art made of record and not relied upon is considered pertinent to applicant's disclosure.

1. "CHEEK RETRACTOR AND MOBILE DEVICE HOLDER" – Meyer et al., US PGPub 2018/0228359 A1.

2. "IMAGING DEVICE FOR DENTAL INSTRUMENTS AND METHODS FOR INTRA-ORAL VIEWING" – Karazivan et al., US PGPub 2012/0040305 A1.

3. "INTRA-ORAL CAMERA" – Matthews, US Pat 9939714 B1.

4. "METHODS AND APPARATUSES FOR DENTAL IMAGES" – Carrier, Jr. et al., US PGPub 2018/0125610 A1.

5. "SOFT HEAD MOUNTED DISPLAY GOGGLES FOR USE WITH MOBILE COMPUTING DEVICES" – Lyons, US PGPub 2015/0234192 A1.

6. "Dental Informatics and Intra-oral Photography in Communicating with Dental Students in the Dominican Republic" - Lawrence PARRISH, Anton DIY, Nicholas R. KENNING, Kristen TEMPLETON, Ruben SAGUN, Nicole S. KIMMES, Gene GASPARD, Stephen J. HESS; Journal of Health Informatics in Developing Countries; April 30, 2014.

7. "A 3-D Reconstruction System for the Human Jaw Using a Sequence of Optical Images" - Sameh M. Yamany, Aly A. Farag, David Tasman, Allan G. Farman; IEEE TRANSACTIONS ON MEDICAL IMAGING, VOL. 19, NO. 5, MAY 2000.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to MAINUL HASAN whose telephone number is (571)272-0422. The examiner can normally be reached on MON-FRI: 10AM-6PM, Alternate FRIDAYS, EST.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, JAY PATEL can be reached on (571)272-2988. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR

system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/Mainul Hasan/
Primary Examiner, Art Unit 2485

EXHIBIT R-11

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
16/951,401	11/18/2020	Philippe SALAH	N&P-51400US2	3930

108676 7590 02/11/2022

Ronald M. Kachmarik
Cooper Legal Group LLC
1388 Ridge Road, Unit 1
Hinckley, OH 44233

EXAMINER	
HASAN, MAINUL	

ART UNIT	PAPER NUMBER
2485	

NOTIFICATION DATE	DELIVERY MODE
02/11/2022	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

docketing@cooperlegalgroup.com

Office Action Summary	Application No. 16/951,401	Applicant(s) SALAH et al.	
	Examiner MAINUL HASAN	Art Unit 2485	AIA (FITF) Status Yes

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTHS FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☑ Responsive to communication(s) filed on 14 January 2022.

☐ A declaration(s)/affidavit(s) under **37 CFR 1.130(b)** was/were filed on _____.

2a)☑ This action is **FINAL.** 2b) ☐ This action is non-final.

3)☐ An election was made by the applicant in response to a restriction requirement set forth during the interview on _____; the restriction requirement and election have been incorporated into this action.

4)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims*

5) ☑ Claim(s) 1-28 is/are pending in the application.

5a) Of the above claim(s) _____ is/are withdrawn from consideration.

6) ☐ Claim(s) _____ is/are allowed.

7) ☑ Claim(s) 1-28 is/are rejected.

8) ☐ Claim(s) _____ is/are objected to.

9) ☐ Claim(s) _____ are subject to restriction and/or election requirement

* If any claims have been determined allowable, you may be eligible to benefit from the **Patent Prosecution Highway** program at a participating intellectual property office for the corresponding application. For more information, please see http://www.uspto.gov/patents/init_events/pph/index.jsp or send an inquiry to **PPHfeedback@uspto.gov.**

Application Papers

10)☐ The specification is objected to by the Examiner.

11)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

Priority under 35 U.S.C. § 119

12)☑ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

Certified copies:

a)☑ All b)☐ Some** c)☐ None of the:

1.☑ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

** See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☑ Notice of References Cited (PTO-892)

2) ☑ Information Disclosure Statement(s) (PTO/SB/08a and/or PTO/SB/08b) Paper No(s)/Mail Date _____.

3) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date _____.

4) ☐ Other: _____.

DETAILED ACTION

Notice of Pre-AIA or AIA Status

The present application, filed on or after March 16, 2013, is being examined under the first inventor to file provisions of the AIA. In the event the determination of the status of the application as subject to AIA 35 U.S.C. 102 and 103 (or as subject to pre-AIA 35 U.S.C. 102 and 103) is incorrect, any correction of the statutory basis for the rejection will not be considered a new ground of rejection if the prior art relied upon, and the rationale supporting the rejection, would be the same under either status.

Response to Amendments

The Applicant's amendments filed on 01/14/2022 have been acknowledged and entered for consideration. No claims have been cancelled nor any new claims added. Claims 1-28 remain pending in the current Application. The Applicant's amendments are in response to the Non-Final Office Action mailed on 10/22/2021.

Claim Rejections - 35 USC § 102

The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

> A person shall be entitled to a patent unless –
>
> (a)(1) the claimed invention was patented, described in a printed publication, or in public use, on sale or otherwise available to the public before the effective filing date of the claimed invention.

Claims 1, 3, 12-19, 21-28 are rejected under AIA 35 U.S.C. 102(a)(1) as being anticipated by Prakash et al. (US PGPub 2013/0209954 A1).

Regarding claim **1** (Original), **Prakash et al.** disclose *a patient-operated imaging device* **(Fig. 5B)** *comprising:*

- *a support* **(Figs. 2A-C,** reference numeral **220. Figs. 5A-B,** reference numeral **520)***;*

- *a mouth retractor formed as an integral part of the support and defining a retractor opening* **(Figs. 2A-B,** reference numeral **210** is the mouthpiece or retractor and **211** is the opening. **Figs. 5A-B,** reference numeral **510** is the mouthpiece or retractor and **511** is the opening. To be specific, reference numerals **212a-b**, known as bite guides, on the mouthpiece **210** act as the mouth retractors as shown in **Fig. 2A)***; and*

- *a mechanism for fastening an image acquisition apparatus to the support in a position in which the acquisition apparatus is oriented so as to receive an image of the retractor opening* **([0062]; Fig. 2B,** reference numeral **226a-c)**,

wherein the support takes the form of a box **(Fig. 5B,** reference numeral **520** shows the box shaped camera mount as the support) *that is in communication with the outside via the retractor opening and via an acquisition opening through which the image acquisition apparatus fastened to the support receives the image of the retractor opening* **([0062]; Fig. 5A** shows the retractor or the mouthpiece **510** and the opening **511** through which the image acquisition device or the cellphone camera receives the image for capturing)*, the support being configured so that the image acquisition apparatus observes the retractor opening regardless of the configuration of the support* **([0069]-[0070]; Figs. 3A-C** show the cellphone camera acquiring the images of the oral cavity through the opening regardless of the configuration of the mount)*;*

the mechanism being chosen from the group consisting of an elastic member, clip-fastening means **([0062]; Fig. 2B,** reference numeral **226a-c)**, *self-gripping strips of hook and loop fastener type, clamping jaws, screws, magnets, and complementarity of shape between the support and the image acquisition apparatus, or consisting of a cover that may be clamped against the support,*

wherein the patient-operated imaging device is adapted to obtain a plurality of images, wherein at least two of the plurality of images correspond to different angles with respect to the patient's teeth **([0069]-[0070]; Figs. 3A-C** show capturing of plurality of images by the cellphone and as the drawings indicate the plurality of images are taken at different angles with respect to the patient's teeth).

Regarding claim **3** (Currently Amended), **Prakash et al.** disclose *an imaging kit* **(Fig. 2C,** reference numeral **250)** *comprising:*

- the patient operated imaging device as claimed in claim 1 **(Fig. 2C,** reference numeral **200)**; *and*

- an image acquisition apparatus **(Fig. 2C,** reference numeral **280)** *that is fastened to the patient operated imaging device in a position in which the image acquisition apparatus is oriented to receive an image of the retractor opening* **([0062]; Fig. 2B,** reference numeral **226a-c)**.

Regarding claim **12** (Currently Amended), **Prakash et al.** disclose *the patient operated imaging device as claimed in claim 1, in which the support defines a closed chamber when the*

opening of the mouth retractor and the acquisition opening are obturated **(Fig. 5B,** reference numeral **520)**.

Regarding claim **13** (Currently Amended), **Prakash et al.** disclose *the patient operated imaging device as claimed in claim 12, in which a lateral wall delimiting the chamber is formed or consists of a material that does not allow the content of the chamber to be accurately discerned* **(Fig. 5B,** reference numeral **520)**.

Regarding claim **14** (Currently Amended), **Prakash et al.** disclose *the patient operated imaging device as claimed in claim 13, in which the lateral wall is opaque, so that an inner volume of the chamber receives substantially no light from outside of the chamber in a service position* **(Fig. 5B,** reference numeral **520)**.

Regarding claim **15** (Currently Amended), **Prakash et al.** disclose *the patient operated imaging device as claimed in claim 1, in which the mouth retractor comprises lobes that are arranged so as to spread cheeks of a patient away from teeth of the patient* **(Fig. 2A,** reference numerals **212a-b** and **Fig. 5A,** reference numerals **512** represent the bite guides which act as a spreader of the patient's cheek as described in **[0004], L3-10)**.

Regarding claim **16** (Currently Amended), **Prakash et al.** disclose *the patient operated imaging device as claimed in claim 1, further comprising a light source that is oriented toward the retractor opening so as to illuminate teeth of a patient through the retractor opening* **(Fig. 2A,** reference numerals **216a-d; Fig. 5,** reference numeral **516)**.

Regarding claim **17** (Currently Amended), **Prakash et al.** disclose *the patient operated imaging device as claimed in claim 16, in which the light source is configured so as to project, through the retractor opening, a reference frame onto the teeth* (**Fig. 5,** reference numeral **516** shows the light source that projects light through the opening **511**. In **[0090], L3-5,** it describes two modes of lighting, e.g. bright field and auto-fluorescent, each of which is equivalent to a reference frame).

Regarding claim **18** (Currently Amended), **Prakash et al.** disclose *the patient operated imaging device as claimed in claim 17, further comprising a monitoring module configured to monitor properties of radiation emitted by the light source as a function of the luminous radiation received by the retractor opening* (**[0037], L9-16;** it discloses detection of auto-fluorescent radiation after illuminating the teeth area with light of a certain wavelength).

Regarding claim **19** (Currently Amended), **Prakash et al.** disclose *the patient operated imaging device as claimed in claim 1, in which the image acquisition apparatus has an objective and is positioned with respect to the acquisition opening so that the objective is maintained substantially in the center of the acquisition opening* (**[0041]; [0058], L5-9;** it teaches placing a lens (e.g. a fish-eye lens) on the optical path as an objective lens).

Regarding claim **21** (Currently Amended), **Prakash et al.** disclose *the patient operated imaging device as claimed in claim 1, in which the retractor opening is configured so that both*

teeth of an upper dental arch of the patient and teeth of a lower arch of the patient are fully visible by the image acquisition apparatus (**Figs. 9A-C**).

Regarding claim **22** (Currently Amended), **Prakash et al.** disclose *the patient operated imaging device as claimed in claim 1, in which the support has a lateral wall which extends between two end faces of the support defining the retractor opening and the acquisition opening, respectively, said lateral wall being rectangular in cross section* (**Fig. 5B** shows the lateral part of the camera mount **520** which is rectangular in cross-section).

Regarding claim **23** (Currently Amended), **Prakash et al.** disclose *the patient operated imaging device as claimed in claim 1, in which the mechanism for fastening the acquisition apparatus is configured so that the acquisition apparatus may be fastened to the support in only one predetermined position* (**[0062];** it discloses that the adjustable clips **226** enable the image acquisition device to be fastened to the mount **220** in different positions, however, it also discloses that in certain scenario, e.g., for a camera cell phone, the clips are not adjustable along a track, meaning the fastening of the camera is achieved only in one position).

Regarding claim **24** (Currently Amended), **Prakash et al.** disclose *the patient operated imaging device as claimed in claim 1, in which the retractor includes a rim extending around the retractor opening and arranged in such a way that the patient's lips may rest on it, leaving the patient's teeth visible through said retractor opening* (**[0102], L3-10; Fig. 5A** shows the rim along the bite guides **512** which allows the patients lips to rest on it while opening the mouth up).

Regarding claim **25** (Currently Amended), **Prakash et al.** disclose *the patient operated imaging device as claimed in claim 24, in which the rim has the shape of a channel configured to hold the patient's lips* **([0102], L3-10; Fig. 5A** shows the rim in the shape of a channel along the bite guides **512** which allows the patients lips to rest on it while opening the mouth up).

Regarding claim **26** (Currently Amended), **Prakash et al.** disclose *the patient operated imaging device as claimed in claim 1, in which the retractor opening is curved around an axis Y which is substantially vertical in a service position* **(Fig. 5A** shows the retractor opening **511,** the four corners of which are curved around an axis perpendicular the plane of the opening).

Regarding claim **27** (Currently Amended), **Prakash et al.** disclose *the patient operated imaging device as claimed in claim 1, in which the retractor opening is larger than the acquisition opening* **(Fig. 2A** shows the retractor opening **211** larger than the optical path **223** through which the image is acquired by the camera).

Regarding claim **28** (Original), **Prakash et al.** disclose *a patient-operated imaging device* **(Fig. 5B)** *comprising:*

- *a support* **(Figs. 2A-C,** reference numeral **220. Figs. 5A-B,** reference numeral **520)**;

- *a mouth retractor formed as an integral part of the support and defining a retractor opening* **(Figs. 2A-B,** reference numeral **210** is the mouthpiece or retractor and **211** is the opening. **Figs. 5A-B,** reference numeral **510** is the mouthpiece or retractor and **511** is the

opening. To be specific, reference numerals **212a-b**, known as bite guides, on the mouthpiece **210** act as the mouth retractors as shown in **Fig. 2A**)*; and*

- an image acquisition apparatus **(Fig. 2C,** reference numeral **280)** *fastened to the support in a position in which the image acquisition apparatus is oriented so as to receive an image of the retractor opening* **([0062]; Fig. 2B,** reference numeral **226a-c**),

wherein the support takes the form of a box **(Fig. 5B,** reference numeral **520** shows the box shaped camera mount as the support) *that is in communication with the outside via the retractor opening and via an acquisition opening through which the image acquisition apparatus fastened to the support receives the image of the retractor opening* **([0062]; Fig. 5A** shows the retractor or the mouthpiece **510** and the opening **511** through which the image acquisition device or the cellphone camera receives the image for capturing)*, the support being configured so that the image acquisition apparatus observes the retractor opening regardless of the configuration of the support* **([0069]-[0070]; Figs. 3A-C** show the cellphone camera acquiring the images of the oral cavity through the opening regardless of the configuration of the mount),

wherein the patient-operated imaging device is adapted to obtain a plurality of images **([0069]-[0070]; Figs. 3A-C** show capturing of plurality of images by the cellphone),

wherein at least two of the plurality of images correspond to different angles with respect to the patient's teeth **([0069]-[0070]; Figs. 3A-C** show capturing of plurality of images by the cellphone and as the drawings indicate the plurality of images are taken at different angles with respect to the patient's teeth).

Claim Rejections - 35 USC § 103

The following is a quotation of 35 U.S.C. 103 which forms the basis for all obviousness rejections set forth in this Office action:

> A patent for a claimed invention may not be obtained, notwithstanding that the claimed invention is not identically disclosed as set forth in section 102, if the differences between the claimed invention and the prior art are such that the claimed invention as a whole would have been obvious before the effective filing date of the claimed invention to a person having ordinary skill in the art to which the claimed invention pertains. Patentability shall not be negated by the manner in which the invention was made.

Claim 2 is rejected under 35 U.S.C. 103 as being unpatentable over Prakash et al. (US PGPub 2013/0209954 A1) in view of Charles (US PGPub 2014/0005484 A1).

Regarding claim **2** (Currently Amended), **Prakash et al.** teach *the patient operated imaging device as claimed in claim 1.*

Although, **Prakash et al.** in **[0054]** teach that the mouthpiece material is magnetic, and in **[0062]**, it teaches that for a camera cell phone, fastening clips are not adjustable along a track, meaning the fastening of the camera is achieved only in one position, but it does not explicitly teach fastening mechanism of the image acquisition device is magnetic.

However, **Charles** teach a system in the same field of endeavor (**Figs. 18, 19, 20A**), where it teaches the fastening mechanism is magnetic (**Charles; [0037], [0263], L13-19**).

It would have been obvious before the effective filing date of the claimed invention to a person having ordinary skill in the art to combine **Prakash et al**'s invention of techniques of capturing intra-oral images with mobile devices to include **Charles**' usage of magnetic fastening, because this can facilitate convenient attachment within a short period of time (**Charles; [0263], L16-23**).

Claims 4-11 are rejected under 35 U.S.C. 103 as being unpatentable over Prakash et al. (US PGPub 2013/0209954 A1) in view of Shanjani et al. (US PGPub 2018/0000563 A1).

Regarding claim **4** (Currently Amended), **Prakash et al.** teach *the imaging kit as claimed in claim 3.*

But, **Prakash et al.** do not teach that *the patient operated imaging device further comprises a detection member and the image acquisition apparatus further comprises a detector that is configured to detect the detection member when the detection member is less than 20 cm from the patient operated imaging device.*

However, **Shanjani et al.** teach a system in the same field of endeavor **(Abstract)**, where it teaches *the patient operated imaging device further comprises a detection member* **(Shanjani et al.; Fig. 37A** shows a phone having a magnetometer which can detect a magnetic field produced by the NFC device (i.e., detection member is the NFC chip in this example)) *and the image acquisition apparatus further comprises a detector that is configured to detect the detection member when the detection member is less than 20 cm from the patient operated imaging device* **(Shanjani et al.; Fig. 37A** shows a phone having a magnetometer which can detect a magnetic field produced by the NFC device which is less than 20cm from the phone **([0163], [0135], [0278], [0014]**); **[0013]**: The removable mechanical activation interrupt may comprise a magnetic switch, a removable activation rod, a pin, etc. Any of these apparatuses may include the dental appliance (e.g., an aligner such as a shell aligner) to which the monitoring apparatus (e.g., ECI) may be permanently or removably coupled; **[0121]**: The one or more proximity sensors may comprise one or more of: a capacitive sensor, an eddy-current sensor, a magnetic sensor; **[0138]**: the sensors herein can be configured as a switch that is activated and/or

deactivated in response to a particular type of signal (e.g., optical, electrical, magnetic, mechanical, etc.); **[0146]**: Alternatively, the intraoral appliance can be operably coupled to a plurality of monitoring devices, such as at least two, three, four, five, or more monitoring devices. Some or all of the monitoring devices may be of the same type (e.g., collect the same type of data). Alternatively, some or all of the monitoring devices may be of different types (e.g., collect different types of data). Any of the embodiments of monitoring devices described herein can be used in combination with other embodiments in a single intraoral appliance).

It would have been obvious before the effective filing date of the claimed invention to a person having ordinary skill in the art to combine **Prakash et al**'s invention of techniques of capturing intra-oral images with mobile devices to include **Shanjani et al**'s detection member, because it creates a device which uses the sensors as a switch that is activated and/or deactivated in response to a particular type of signal (e.g., optical, electrical, magnetic, mechanical, etc.) **(Shanjani et al.; [0138])**.

Regarding claim **5** (Currently Amended), **Prakash et al.** and **Shanjani et al.** teach *the imaging kit as claimed in claim 4, in which the detector is configured so as to detect the detection member only when the detection member is less than 20 cm from the patient operated imaging device* **(Shanjani et al.; [0163], [0135], [0278], [0014]; Fig. 37A** shows a phone having a magnetometer which can detect a magnetic field produced by the NFC device which is less than 20cm from the phone).

Regarding claim **6** (Original), **Prakash et al.** and **Shanjani et al.** teach *the imaging kit as claimed in claim 4, in which the detection member is positioned less than 5 cm from the edge of*

the acquisition opening (**Shanjani et al.; [0163], [0135], [0278], [0014]; Fig. 37A** shows a phone having a magnetometer which can detect a magnetic field produced by the NFC device which is less than 5cm from the phone).

Regarding claim **7** (Original), **Prakash et al.** and **Shanjani et al.** teach *the imaging kit as claimed in claim 4, in which the detector further comprises a magnetometer* (**Shanjani et al.; [0048], [0121], [0138]**).

Regarding claim **8** (Original), **Prakash et al.** and **Shanjani et al.** teach *the imaging kit as claimed in claim 4, in which the detector is configured to trigger an execution of a computer program loaded on a processing module of the image acquisition apparatus in the event that the detection member is detected* (**Shanjani et al.**; **[0145]**: In some embodiments, the monitoring device **300** is dormant before being delivered to the patient (e.g., during storage, shipment, etc.) and is activated only when ready for use. This approach can be beneficial in conserving power expenditure. For example, the components of the monitoring device **300** can be electrically coupled to the power source **316** at assembly, but may be in a dormant state until activated, e.g., by an external device such as a mobile device, personal computer, laptop, tablet, wearable device, power hub etc. The external device can transmit a signal to the monitoring device **300** that causes the activation mechanism to activate the monitoring device **300**; **[0258]**: As mentioned above, any of the apparatuses described herein (including systems) may communicate with a hand-held electronic device such a smartphone via control software running on the smartphone (or other hand-held electronics). This application software may interface with the electronic compliance indicator and may enhance wireless communications between the

electronic compliance indicator (ECI) using NFC and BLE protocols… An ECI apparatus may generally record sensor data from patients wearing an orthodontic appliance such as an aligner. The data may be stored in physical memory on the ECI and retrieved by another device, e.g., using NFC and BLE technologies as described above (or NFC and NFC), so that the smartphone may retrieve the data. The smartphone application (app) may consist of several components, some of which are described in **FIGS. 41, 42** and **43**. For example, in **FIG. 41** schematically illustrates an NFC/BLE communication control. In addition, **FIGS. 44, 45** and **46** schematically illustrate operational states of the ECI device, as well as control of communication between the device and a remote processor (e.g., smartphone)).

It would have been obvious before the effective filing date of the claimed invention to a person having ordinary skill in the art to combine **Prakash et al**'s invention of techniques of capturing intra-oral images with mobile devices to include **Shanjani et al**'s triggering of computer program, because it conserves power **(Shanjani et al.; [0145]**).

Regarding claim **9** (Original), **Prakash et al.** and **Shanjani et al.** teach *the imaging kit as claimed in claim 8, in which the processing module is configured to acquire, in response to the detection of the detection member by the detector, one or more updated images, then analyze the one or more updated images* **(Shanjani et al.; [0142], L8-9)** *to detect an incorrect positioning of the image acquisition apparatus and/or an incorrect positioning of the retractor* **(Shanjani et al.; [0249], L18-26)** *and/or a poor illumination of the retractor opening and/or an unsuitable support length* **(Shanjani et al.; [0145]**: In some embodiments, the monitoring device **300** is dormant before being delivered to the patient (e.g., during storage, shipment, etc.) and is activated only when ready for use. This approach can be beneficial in conserving power

expenditure. For example, the components of the monitoring device **300** can be electrically coupled to the power source **316** at assembly, but may be in a dormant state until activated, e.g., by an external device such as a mobile device, personal computer, laptop, tablet, wearable device, power hub etc. The external device can transmit a signal to the monitoring device **300** that causes the activation mechanism to activate the monitoring device **300**; **[0164]**: Some of the proximity sensor types described herein (e.g., capacitive sensors) may also be touch sensors, such that they are activated both by proximity to the sensing target as well as direct contact with the target; **[0171]**: Although **FIG. 8B** illustrates a single monitoring device **850** with a single capacitive sensor **854**, other configurations can also be used. For example, in alternative embodiments, the monitoring device **850** can include multiple capacitive sensors located at different sites on the appliance **852** to detect proximity to and/or contact with multiple locations in the intraoral cavity. Optionally, multiple monitoring devices can be used, with each device being coupled to one or more respective capacitive sensors).

It would have been obvious before the effective filing date of the claimed invention to a person having ordinary skill in the art to combine **Prakash et al**'s invention of techniques of capturing intra-oral images with mobile devices to include **Shanjani et al**'s triggering of computer program, because it conserves power **(Shanjani et al.; [0145])**.

Regarding claim **10** (Currently Amended), **Prakash et al.** and **Shanjani et al.** teach *the imaging kit as claimed in claim 4, in which the image acquisition apparatus transmits a message in response to the detection of the detection member by the detector, the message relating to the use of the patient operated imaging device and/or relating to the fastening of the acquisition apparatus and/or relating to the fastening of the dental retractor and/or relating to the timing of*

the updated images to be acquired **(Shanjani et al.; [0145]**: In some embodiments, the monitoring device **300** is dormant before being delivered to the patient (e.g., during storage, shipment, etc.) and is activated only when ready for use. This approach can be beneficial in conserving power expenditure. For example, the components of the monitoring device **300** can be electrically coupled to the power source **316** at assembly, but may be in a dormant state until activated, e.g., by an external device such as a mobile device, personal computer, laptop, tablet, wearable device, power hub etc. The external device can transmit a signal to the monitoring device **300** that causes the activation mechanism to activate the monitoring device **300)**.

It would have been obvious before the effective filing date of the claimed invention to a person having ordinary skill in the art to combine **Prakash et al**'s invention of techniques of capturing intra-oral images with mobile devices to include **Shanjani et al**'s detection member, because it creates a device which uses the sensors as a switch that is activated and/or deactivated in response to a particular type of signal (e.g., optical, electrical, magnetic, mechanical, etc.) **(Shanjani et al.; [0138]**).

Regarding claim **11** (Currently Amended), **Prakash et al.** teach *the patient operated imaging device as claimed in claim 1.*

But it does not explicitly teach that *the mechanism for fastening includes the detection member.*

However, **Shanjani et al.** teach a system in the same field of endeavor **(Abstract)**, where it teaches that *the mechanism for fastening includes the detection member* **(Shanjani et al.; [0148]**).

It would have been obvious before the effective filing date of the claimed invention to a person having ordinary skill in the art to combine **Prakash et al**'s invention of techniques of capturing intra-oral images with mobile devices to include **Shanjani et al**'s usage of the detection member on the fastener, because it is beneficial to distribute the components of the monitoring device across multiple appliances in order to accommodate space limitations, accommodate power limitations, and/or improve sensing **(Shanjani et al.; [0152], L12-16)**.

Claim 20 is rejected under 35 U.S.C. 103 as being unpatentable over Prakash et al. (US PGPub 2013/0209954 A1) in view of Pfeiffer (US Pat 5,677,537).

Regarding claim **20** (Currently Amended), **Prakash et al.** teach *the patient operated imaging device as claimed in claim 1.*

But **Prakash et al.** do not explicitly teach the mechanism of an elastic member.

However, **Pfeiffer** teaches a system in the same field of endeavor **(Abstract)**, where it teaches the fastening mechanism is made of an elastic member **(Pfeiffer; Fig. 2, Col 3, L40-45)**.

It would have been obvious before the effective filing date of the claimed invention to a person having ordinary skill in the art to combine **Prakash et al**'s invention of techniques of capturing intra-oral images with mobile devices to include **Pfeiffer**'s usage of elastic material, because the holder (retractor) and sensor (camera) are secured in a defined alignment relative to one another **(Pfeiffer; Fig. 2, Col 4, L4-8)**.

Response to Arguments

Applicant's arguments filed on 01/14/2022 have been fully considered but they are not persuasive.

The Applicant's filing of the Terminal Disclosure dated 01/14/2022 and subsequent approval of the TD on 01/14/2022 has rendered the previously set forth Double Patenting rejection moot and therefore withdrawn.

The Applicant's amendments to Abstract have appropriately addressed the previously set forth objection and therefore withdrawn.

The Applicant's amendments made in the claims have appropriately addressed the previously set forth 112(b) indefiniteness rejection and therefore withdrawn.

The Applicant in P9 of the remark section argues regarding the rejection of claim 1 by **Prakash et al. (US PGPub 2013/0209954 A1)** by stating that "PRAKASH discloses a device comprising a mouthpiece including an upper bite guide and a lower bite guide to acquire images of the oral cavity of a subject. PRAKASH does not deal with a "mouth retractor', to push the lips away from the teeth, so as to expose the teeth to the imaging device. On the contrary, the device of PRAKASH causes a subject to bite against the bite guides of the mouthpiece. PRAKASH insists on the need for the upper and lower guides and for the patient to bite the upper and lower guides to expose the subject's oral cavity (see for example paragraphs [0004], [0009], [0048], [0078], [0103] and [0106]). More generally, the aim of PRAKASH is to acquire images of the

oral cavity. The device of PRAKASH does not allow the acquisition of the front teeth of a patient (see FIG. 9). Indeed, the device of PRAKASH hides at least part of teeth of the subject as the gums or teeth of the subject's upper jaw are placed inside the upper bite guide 212a (see [0048))".

The Examiner cannot concur with the Applicant and respectfully disagrees. The Applicant argues that **Prakash et al.** do not teach "mouth retractor". However, as per dictionary.com definition, a retractor is "an instrument or appliance for drawing back an impeding part". **Prakash et al.** in **Fig. 5D** shows a mouthpiece **560** whose shape is rectangular in nature with an opening in the middle. As shown in the drawing the mouthpiece opens up the mouth cavity so that a camera can take a picture through the opening. Therefore, the mouthpiece in the drawing is functioning as a mouth retractor which is also shown in **Fig. 5B** with **510**. Moreover, the bite guide **512**, as part of the mouthpiece **510,** which are shown in **Fig. 5A,** forces the mouth cavity or lips of the patient to retract or open wide. Therefore, for all practical purpose the mouthpiece mechanism with the bite guide is indeed a mouth retractor. The Applicant also argues that **Prakash et al**'s device captures the image of the oral cavity, not the front teeth of a patient. However, the drawings of the images captured by **Prakash et al**'s device in **Figs. 8, 9** clearly show the image of the teeth of a patient. The limitations do not claim that the patient operated imaging device is capturing images of a patient's front teeth. *In arguendo*, even if the claim recites that the device is meant for capturing images of the front teeth, **Prakash et al.** still would have anticipated the limitation because capturing images of the front teeth would have been an intended use of the device and therefore would not have carried any patentable weight. Therefore, the Examiner believes the amended limitations of claim(s) are still anticipated by **Prakash et al.**

Conclusion

THIS ACTION IS MADE FINAL. Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the mailing date of this final action.

The prior art made of record and not relied upon is considered pertinent to applicant's disclosure.

1. "INTRA-ORAL CAMERA" – Matthews, US Pat 9939714 B1.

2. "METHODS AND APPARATUSES FOR DENTAL IMAGES" – Carrier, Jr. et al., US PGPub 2018/0125610 A1.

3. "SOFT HEAD MOUNTED DISPLAY GOGGLES FOR USE WITH MOBILE COMPUTING DEVICES" – Lyons, US PGPub 2015/0234192 A1.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to MAINUL HASAN whose telephone number is (571)272-0422. The examiner can normally be reached on MON-FRI: 10AM-6PM, Alternate FRIDAYS, EST.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, JAY PATEL can be reached on (571)272-2988. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/Mainul Hasan/
Primary Examiner, Art Unit 2485

EXHIBIT R-12

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Appl. No.	:	16/951,401
Applicant	:	Philippe SALAH
Filed	:	November 18, 2020
Title	:	DENTAL IMAGING DEVICE
Conf. No.	:	3930
TC/A.U.	:	2485
Examiner	:	Mainul Hasan
Customer No.	:	108676
Docket No	:	N&P-51400US2

Mail Stop Amendment
Commissioner for Patents
P.O. Box 1450
Alexandria VA 22313-1450

Amendment "G" After Final

Sir:

This Amendment is in response to the Office Action dated February 11, 2022, Paper No./Mail Date: 20220129. The three-month period for responding to the Office Action expired on May 11, 2022. Accordingly, the Applicant respectfully requests and petitions that the response date be extended for two months, up to and including July 11, 2022. The $640.00 two-month extension of time fee is being paid via credit card with the filing of this amendment.

Amendments to the Claims are reflected in the listing of claims that begins on page 2 of this paper.

Remarks / Arguments begin on page 7 of this paper.

Listing of Claims:

The listing of claims will replace all prior versions, and listings, of claims in this application:

1. (Currently amended) A patient-operated imaging device comprising:
   - a support;
   - a mouth retractor formed as an integral part of the support and defining a retractor opening; and
   - a mechanism for fastening an image acquisition apparatus to the support in a position in which the acquisition apparatus is oriented so as to receive an image of the retractor opening through which, in a service position, front teeth of the patient are visible,

   wherein the support takes the form of a box that is in communication with the outside via the retractor opening and via an acquisition opening through which the image acquisition apparatus fastened to the support receives the image of the retractor opening, the support being configured so that the image acquisition apparatus observes the retractor opening regardless of the configuration of the support;

   the mechanism being chosen from the group consisting of an elastic member, clip-fastening means, self-gripping strips of hook and loop fastener type, clamping jaws, screws, magnets, and complementarity of shape between the support and the image acquisition apparatus, or consisting of a cover that may be clamped against the support,

   wherein the patient-operated imaging device is adapted to obtain a plurality of images, wherein at least two of the plurality of images correspond to different angles with respect to the patient's teeth.

2. (Previously amended) The patient-operated imaging device as claimed in claim 1, in which the mechanism is magnetic and configured so that the image acquisition apparatus may be fastened to the support in only one predetermined position.

3. (Previously amended) An imaging kit comprising:

- the patient-operated imaging device as claimed in claim 1; and

- an image acquisition apparatus that is fastened to the patient-operated imaging device in a position in which the image acquisition apparatus is oriented to receive an image of the retractor opening.

4. (Previously amended) The imaging kit as claimed in claim 3, in which the patient-operated imaging device further comprises a detection member and the image acquisition apparatus further comprises a detector that is configured to detect the detection member when the detection member is less than 20 cm from the patient-operated imaging device.

5. (Previously amended) The imaging kit as claimed in claim 4, in which the detector is configured so as to detect the detection member only when the detection member is less than 20 cm from the patient-operated imaging device.

6. (Original) The imaging kit as claimed in claim 4, in which the detection member is positioned less than 5 cm from the edge of the acquisition opening.

7. (Original) The imaging kit as claimed in claim 4, in which the detector further comprises a magnetometer.

8. (Original) The imaging kit as claimed in claim 4, in which the detector is configured to trigger an execution of a computer program loaded on a processing module of the image acquisition apparatus in the event that the detection member is detected.

9. (Original) The imaging kit as claimed in claim 8, in which the processing module is configured to acquire, in response to the detection of the detection member by the detector, one or more updated images, then analyze the one or more updated images to detect an incorrect positioning of the image acquisition apparatus and/or an incorrect positioning of the retractor and/or a poor illumination of the retractor opening and/or an unsuitable support length.

10. (Previously amended) The imaging kit as claimed in claim 4, in which the image acquisition apparatus transmits a message in response to the detection of the detection member by the detector, the message relating to the use of the patient-operated imaging device and/or relating to the fastening of the acquisition apparatus and/or relating to the fastening of the dental retractor and/or relating to the timing of the updated images to be acquired.

11. (Previously amended) The patient-operated imaging device as claimed in claim 1, in which the mechanism for fastening includes the detection member.

12. (Previously amended) The patient-operated imaging device as claimed in claim 1, in which the support defines a closed chamber when the opening of the mouth retractor and the acquisition opening are obturated.

13. (Previously amended) The patient-operated imaging device as claimed in claim 12, in which a lateral wall delimiting the chamber is formed or consists of a material that does not allow the content of the chamber to be accurately discerned.

14. (Previously amended) The patient-operated imaging device as claimed in claim 13, in which the lateral wall is opaque, so that an inner volume of the chamber receives no light from outside of the chamber in a service position.

15. (Previously amended) The patient-operated imaging device as claimed in claim 1, in which the mouth retractor comprises lobes that are arranged so as to spread cheeks of a patient away from teeth of the patient.

16. (Previously amended) The patient-operated imaging device as claimed in claim 1, further comprising a light source that is oriented toward the retractor opening so as to illuminate teeth of a patient through the retractor opening.

17. (Previously amended) The patient-operated imaging device as claimed in claim 16, in which the light source is configured so as to project, through the retractor opening, a reference frame onto the teeth.

18. (Previously amended) The patient-operated imaging device as claimed in claim 17, further comprising a monitoring module configured to monitor properties of radiation emitted by the light source as a function of the luminous radiation received by the retractor opening.

19. (Previously amended) The patient-operated imaging device as claimed in claim 1, in which the image acquisition apparatus has an objective and is positioned with respect to the acquisition opening so that the objective is maintained in the center of the acquisition opening.

20. (Previously amended) The patient-operated imaging device as claimed in claim 1, in which the mechanism is an elastic member.

21. (Previously amended) The patient-operated imaging device as claimed in claim 1, in which the rectractor opening is configured so that both teeth of an upper dental arch of the patient and teeth of an lower arch of the patient are fully visible by the image acquisition apparatus.

22. (Previously amended) The patient-operated imaging device as claimed in claim 1, in which the support has a lateral wall which extends between two end faces of the support defining the retractor opening and the acquisition opening, respectively, said lateral wall being rectangular in cross section.

23. (Previously amended) The patient-operated imaging device as claimed in claim 1, in which the mechanism for fastening the acquisition apparatus is configured so that the acquisition apparatus may be fastened to the support in only one predetermined position.

24. (Previously amended) The patient-operated imaging device as claimed in claim 1, in which the retractor includes a rim extending around the retractor opening and arranged in such a way that the patient's lips may rest on it, leaving the patient's teeth visible through said retractor opening.

25. (Previously amended) The patient-operated imaging device as claimed in claim 24, in which the rim has the shape of a channel configured to hold the patient's lips.

26. (Previously amended) The patient-operated imaging device as claimed in claim 1, in which the retractor opening is curved around an axis Y which is vertical in a service position.

27. (Previously amended) The patient-operated imaging device as claimed in claim 1, in which the retractor opening is larger than the acquisition opening.

28. (Original) An patient-operated imaging device comprising:
- a support;
- a mouth retractor formed as an integral part of the support and defining a retractor opening; and
- an image acquisition apparatus fastened to the support in a position in which the image acquisition apparatus is oriented so as to receive an image of the retractor opening,

wherein the support takes the form of a box that is in communication with the outside via the retractor opening and via an acquisition opening through which the image acquisition apparatus fastened to the support receives the image of the retractor opening, the support being configured so that the image acquisition apparatus observes the retractor opening regardless of the configuration of the support,

wherein the patient-operated imaging device is adapted to obtain a plurality of images, wherein at least two of the plurality of images correspond to different angles with respect to the patient's teeth.

REMARKS

Reconsideration of the subject application in view of the present Amendment is respectfully requested.

By the present amendment, claim 1 is amended. The amendment is in view of Fig. 3.

With regard to the rejections based upon PRAKASH (US 2013/0209954), please note the following:

Claim 1 recites:

> "*A patient-operated imaging device comprising:*
> *-a support;*
> *-a mouth retractor formed as an integral part of the support and defining a retractor opening; and*
> *-a mechanism for fastening an image acquisition apparatus to the support in a position in which the acquisition apparatus is oriented so as to receive an image of the retractor opening through which, in a service position, front teeth of the patient are visible,*
> *wherein the support takes the form of a box that is in communication with the outside via the retractor opening and via an acquisition opening through which the image acquisition apparatus fastened to the support receives the image of the retractor opening, the support being configured so that the image acquisition apparatus observes the retractor opening regardless of the configuration of the support;*
> *the mechanism being chosen from the group consisting of an elastic member, clip-fastening means, self-gripping strips of hook and loop fastener type, clamping jaws, screws, magnets, and complementarity of shape between the support and the image acquisition apparatus, or consisting of a cover that may be clamped against the support,*
> *wherein the patient-operated imaging device is adapted to obtain a plurality of images, wherein at least two of the plurality of images correspond to different angles with respect to the patient's teeth."*

PRAKASH discloses a device comprising a mouthpiece including an upper bite guide and a lower bite guide to acquire images of the oral cavity of a subject.

Even if some images of oral cavity may include teeth of the subject as shown in figures 8 and 9A-9F of PRAKASH, PRAKASH never suggests acquiring images of the

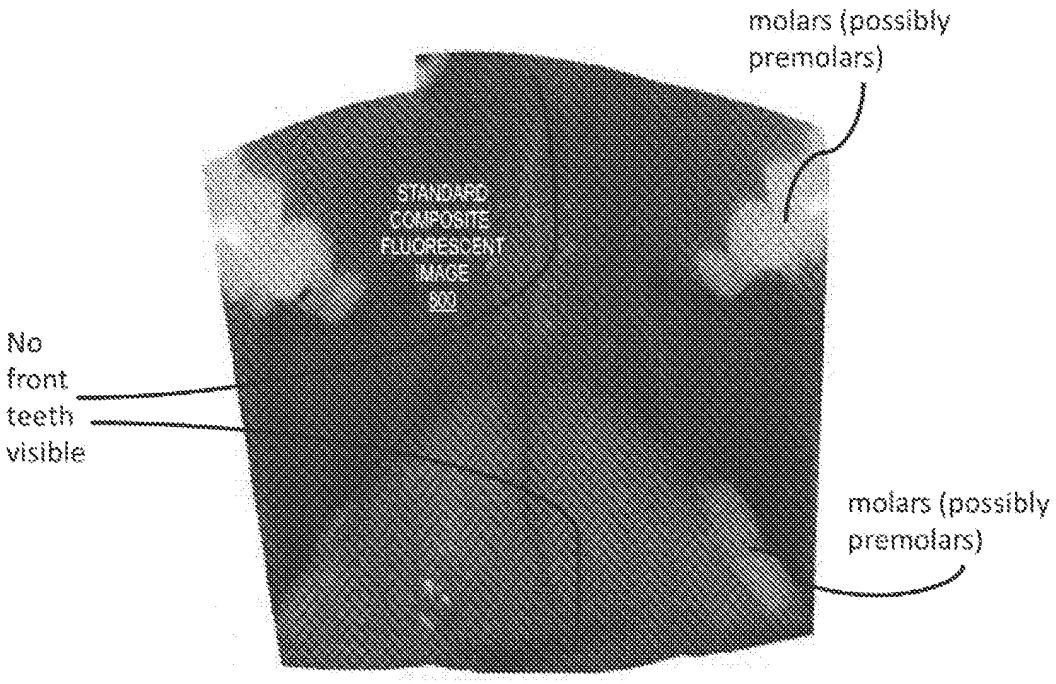
teeth of the subject and above all PRAKASH does not suggest acquiring images of the front teeth of the subject.

Moreover, such acquisition is incompatible with the teaching of PRAKASH. Indeed, PRAKASH does not deal with a device allowing to push the lips away from the teeth, so as to expose the teeth to the imaging device, and even less to push the lips away from the front teeth to expose the front teeth to the imaging device.

On the contrary, the device of PRAKASH causes a subject to bite against the bite guides of the mouthpiece.

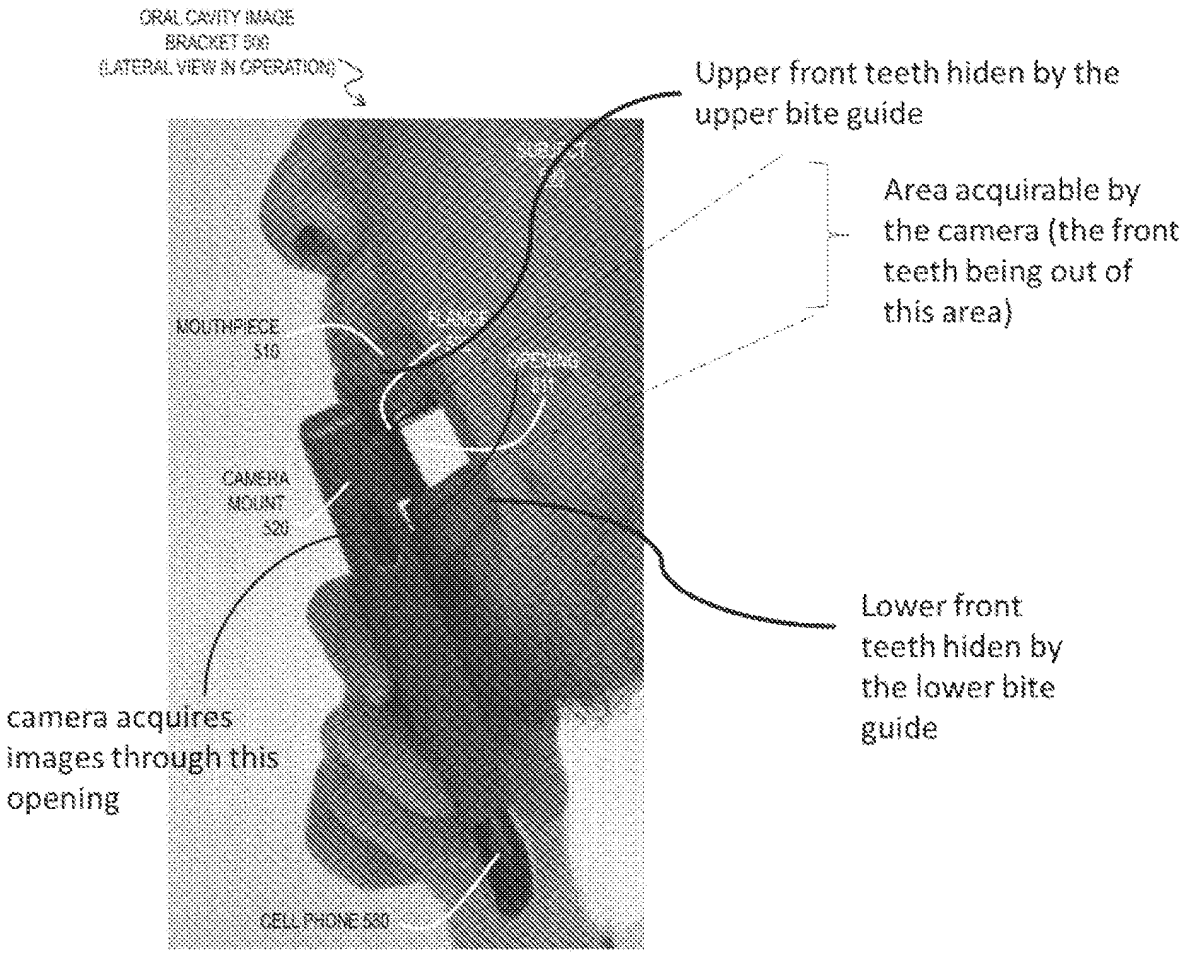
As such, when the subject bites the bite guides in order to acquire images of the oral cavity of the subject, the front teeth are under the lips and out of the view of the imaging device, as illustrated below on the annotated figures 5B and 8 of PRAKASH.

FIG. 8



Annotated fig. 8 of PRAKASH

FIG. 5B



Annotated fig. 5B of PRAKASH

In particular, as specified in paragraph 50 of PRAKASH, the bite guides are separated by an opening that provides a view into the subject's oral cavity. However, as it clearly appears on the annotated figure 5B above, front teeth are not visible thought the opening as they are positioned below and under the opening and separated from this opening by the bite guides.

Furthermore, as shown in figures 9A-9F of PRAKASH, no matter what angle the acquisition device is at, none of them can acquire images of front teeth of the subject.

More generally, the aim of PRAKASH is to acquire images of the oral cavity. There is no indication in PRAKASH that the person of ordinary skill in the art would

have modified the PRAKASH device to achieve the present invention. In particular, acquiring images of the teeth requires experience and skill. Thus, it is not obvious how modify a device configured to acquire images of the oral cavity into a device configured to acquired images of the teeth. As such, the teaching of PRAKASH is not compatible with the invention as claimed in claim 1.

In particular, just because it seems technically simple, a posteriori, does not mean that the development of the invention did not require inventive step.

In conclusion, PRAKASH does not describe the invention as claimed in claim 1, and there is no hint to lead a person of ordinary skilled in the art to the invention as claimed.

Therefore, claim 1, like all claims 2 to 28, is new and inventive.

Accordingly, it is respectfully requested that the rejections based upon PRAKASH (i.e., under 35 U.S.C. §102 and §103) be withdrawn.

In light of the foregoing, it is respectfully submitted that the present application is in condition for allowance and notice to that effect is hereby requested. If it is determined that the application is not in condition for allowance, the Examiner is invited to initiate a telephone interview with the undersigned attorney to expedite prosecution of the present application.

If there are any fees resulting from this communication, please charge same to our Deposit Account No. 505088, our Order No. N&P-51400US2.

Respectfully submitted,
Cooper Legal Group

By: /Ronald M. Kachmarik/
Ronald M. Kachmarik, Reg. No. 34512

1388 Ridge Road
Unit 1
Hinckley, OH 44233
(216) 654-0090